UNIVERSITY OF HELSINKI

DEPARTMENT OF MATHEMATICS AND STATISTICS

MASTER'S THESIS

Existence and Uniqueness of Solutions for Stochastic Differential Equations

Author
Patrik NUMMI

Supervisor
Dr. Lauri VIITASAARI

November 8, 2019

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta/Osasto — Fakultet/Sektion — Faculty	Laitos — Institution — Department
Faculty of Science	Department of Mathematics and Statistics

Tekijä — Författare — Author

Patrik Nummi

Työn nimi — Arbetets titel — Title

Existence and Uniqueness of Solutions for Stochastic Differential Equations

Oppiaine — Läroämne — Subject

Mathematics

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Master's Thesis	November 2019	61 pp.

Tiivistelmä — Referat — Abstract

Stochastic differential equations arise typically in situations where, for instance, the time evolution of a given quantity has some degree of inherent uncertainty. Dating back to Albert Einstein's work in 1905, stochastic differential equations are widely used in applications such as mathematical physics and financial mathematics. Classical examples include the Black-Scholes model, and Ornstein-Uhlenbeck process as the solution of the Langevin equation. In addition, stochastic differential equations have connections to the theory of deterministic partial differential equations, and the Sobolev space theory of deterministic calculus has its counterpart in the stochastic case as well, leading to the so-called Malliavin calculus, or stochastic calculus of variations. There also exists a considerable research literature of stochastic analysis with respect to other processes than Brownian motion, such as Lévy processes.

In this thesis we present an existence and uniqueness theorem for stochastic differential equations with respect to a Brownian motion, under the assumption that the coefficients satisfy Lipschitz and linear growth estimates. The theorem is originally due to Kiyosi Itô. In addition, we present a proof of the continuity of the solution with respect to the initial data, assuming it is deterministic. This theorem was originally proved by Tsukasa Fujiwara and Hiroshi Kunita.

Avainsanat — Nyckelord — Keywords

stochastic differential equations, stochastic analysis, stochastic integration

Säilytyspaikka — Förvaringsställe — Where deposited

Kumpula Campus Library

Muita tietoja — Övriga uppgifter — Additional information

Contents

Chapter 1

Introduction

When microscopic objects are suspended in a fluid, their inherent thermal motion produces collisions with neighbouring objects, resulting in an erratic motion, a random walk. This phenomenon was first brought to greater attention by botanist Robert Brown in 1828, and hence named in his honour as Brownian motion, but it was only in 1905 that Albert Einstein published the first quantitative explanation to this random motion. Subsequent descriptions for this Brownian motion were given by Marian Smoluchowski (1906) and Paul Langevin (1908). Each of these physical theories relied on novel use of established physics, and indeed to some extent played their part in creating new, as Einstein's theory offered further proof of the atomic and molecular structure of matter. For the experiments which confirmed Einstein's theory, physicist Jean Perrin was awarded the Nobel Prize for Physics in 1926.

The first purely mathematical description of Brownian motion was given by Louis Bachelier in 1900, although by a quirk of history his work at the time went unnoticed, and was rediscovered and appreciated a half-century later. Norbert Wiener in his 1923 article [39] gave the first notable mathematical formulation of Brownian motion as a stochastic process. For this reason the mathematical Brownian motion is also referred to as Wiener process. Later Andrey Kolmogorov and Paul Lévy [23], in 1948, also offered their constructions of the Brownian motion.

Einstein derived his results governing Brownian motion from statistical mechanics, and related the probability density of a single Brownian particle to the diffusion equation. Alternatively, for instance Ornstein and Uhlenbeck [38], following Langevin, proposed a different analysis of the dynamics of the Brownian motion. Denoting the position of the particle by $X(t)$, from Newton's second law one can derive the deterministic ordinary differential equation describing the evolution of the position $X(t)$ as a function of time:

$$X'(t) = f(t, X(t)), \quad X(0) = X_0,$$

with f being some sufficiently smooth function describing the particle's speed. Now in order to account for the random perturbation (which may depend on the position of the particle) we add an additional term, so that over the arbitrarily small interval Δt our differential equation becomes

$$dX_t = f(t, X_t)dt + \sigma(t, X_t)dW_t, \quad X(0) = X_0, \tag{1.1}$$

with σ representing the intensity of the random noise and dW_t being, at least in some heuristic sense, the increment of the Brownian motion over Δt. The next question is of the precise mathematical meaning of the above equation. In the theory of ordinary differential equations it is typical to convert an equation such as (1.1) to an equivalent integral equation:

$$X_t = X_0 + \int_0^t f(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, \tag{1.2}$$

but this alone offers no insight, as now the problem is to make sense of the latter integral term in (1.2). The paths of the Brownian motion W_t are almost surely non-differentiable on intervals $[a, b] \subset \mathbb{R}$, a fact first established by Paley, Wiener and Zygmund [33]. Also, the total variation of the map $t \mapsto W_t$ is almost surely non-finite on any $[0, t] \subset \mathbb{R}$. Hence Riemann-Stieltjes integration is out of the question. Kiyosi Itô in his 1944 paper [14] presented was a different integral altogether, later named in his honour. This stochastic integral was by construction a limit of simple integrands, but the limit was taken in L^2. Itô showed that assuming sufficient regularity conditions for the integrand, such integration is indeed possible. It was also Itô, who in his 1946 article [15], attacked the issue of existence and uniqueness of a solution to a problem defined by (1.2).

The existence and uniqueness of solutions of equations such as (1.2) is the subject of this thesis. Indeed, specifically we are considering a stochastic process X solving the following equation:

$$X_t = \xi + \int_0^t \sigma(s, X_s)dW_s + \int_0^t f(s, X_s)ds \tag{1.3}$$

almost surely for all $t \in \mathbb{R}_+$, with ξ a random variable.

Stochastic differential equations of the kind in (1.3) have been widely utilized in physical sciences due to the obvious historical context, but also in financial mathematics, filtering theory, mathematical biology and systems analysis (see [17] and the references therein). Later developments have been related, again much due to historical context, to diffusions [37], linking the theory of stochastic differential equations and partial differential equations, [2]. There exists a considerable body of literature on the

properties of solutions, for instance on weak solutions, and Markov properties, strong and weak. See for instance [17] or [34]. From the 1960's onwards the stochastic integral pioneered by Itô has been generalized to other processes than Brownian motion, for instance to the case of continuous semimartingales. See [20] or [27], or [32] for a more recent overview. In addition, the stochastic integration theory has been extended to Banach-valued processes as well, see for instance [13] and [26].
By extension, also the SDE existence and uniqueness results have been investigated in a more general setting, for instance by Gyöngy & Krylov [11] for semimartingales and Mandrekar & Rüdiger [26] for the Banach-valued case. From the 1960's onwards there have been considerations on pathwise stochastic integration as well, initially with respect to the Brownian motion as in for instance in Wong & Zakai [40], and for semimartingales in Nutz [29]. In addition, the theory of stochastic differential equations has been complemented by backward stochastic differential equations, see [43]. Even more recent development is the pathwise integration using rough paths. First pioneered by Terry Lyons in the 1990's, for instance in [25], it has been employed by Martin Hairer to solve the KPZ equation [12], for which he received the Fields medal in 2014.
In order to present the theorem guaranteeing the solution to (1.3), we need some preliminary tools of probability theory, and these will be presented in Chapter 2. As mentioned above, the main theoretical development in solving the above equation was the Itô integral itself, and its construction along with some key properties will be given in Chapter 3. Although we are primarily interested in integration with respect to the Brownian motion W, the presentation is somewhat more general than necessary and could be accommodated for other stochastic processes as well. Finally, the proof for the existence and uniqueness of solution to (1.3) will be presented in Chapter 4, close in spirit to Itô's original proof, meaning the coefficients $\sigma(t, x)$ and $f(t, x)$ will be assumed to be Lipschitz functions in x. Then, following Hiroshi Kunita's [19] work, we will present his proof of the continuity of the solution to (1.3) with respect to the initial data ξ, assuming ξ is deterministic.

Chapter 2

Preliminaries of Stochastic Analysis

2.1 Stochastic processes, filtrations and stopping times

We assume that $(\Omega, \mathcal{F}, P)$ is a probability space throughout: that is, for a set Ω, the collection $\mathcal{F} \subset 2^{\Omega}$ contains Ω, is closed under complements, and is closed under countable unions. Sets $A \in \mathcal{F}$ are called events. We also have a σ-additive set function $P : \mathcal{F} \to [0, \infty]$ for which $P(\emptyset) = 0$, and $P(\Omega) = 1$. A null set is a set $N \in \mathcal{F}$ such that $P(N) = 0$. A probability space is called *complete* if every subset N_0 of a null set N is an event, and we have $P(N_0) = 0$. A σ-subalgebra $\mathcal{G} \subset \mathcal{F}$ is said to be *augmented* if it contains all the null sets in $(\Omega, \mathcal{F}, P)$.

Definition 2.1. Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $D \subset \mathbb{R}$. A *stochastic process* is a function $X : \mathbb{T} \times \Omega \to \mathbb{R}^d$ such that the mapping $\omega \mapsto X(t, \omega) =: X_t(\omega)$ is a random variable. The function $t \mapsto X_t(\omega)$ is called a *sample path*.

Thus a stochastic process X is a collection of random variables $\{X_t : t \in \mathbb{T}\}$ indexed by a set $\mathbb{T}$. In our case, we will typically take $\mathbb{T}$ to be either $[0, \infty)$ or $[a, b]$, where $a, b \in \mathbb{R}$. Given that a stochastic process is defined on the product space $\mathbb{T} \times \Omega$ with the product measure $dt \otimes dP$, we have several alternatives to speak of the similarity of two stochastic processes:

Definition 2.2. Suppose X and Y are two stochastic processes defined on $(\Omega, \mathcal{F}, \Omega)$. Then Y is said to be a *modification* of X if for every $t \in \mathbb{T}$, $X_t = Y_t$ almost surely. Furthermore, we say X and Y are *indistinguishable* if, for almost every $\omega \in \Omega$, $X_t = Y_t$ for every $t \in \mathbb{T}$.

It is clear that for X and Y to be indistinguishable, they must also be modifications of one another.

There are equally well different alternatives to think of continuity of a stochastic process, given that our process is defined on the product space $\mathbb{T} \times \Omega$. Here we make the standard choice to always restrict ourselves to the continuity of the sample paths in particular:

Definition 2.3. A stochastic process X on a probability space $(\Omega, \mathcal{F}, P)$ is said to be continuous (right/left-continuous) if the function $t \mapsto X_t(\omega)$ is continuous (right/left-continuous) for every $\omega \in \Omega$. A such process is almost surely continuous if the preceding holds for almost every $\omega \in \Omega$.

The next lemma is a useful tool to guarantee sufficient conditions for indistinguishability of two processes:

Lemma 2.4. *Suppose that a stochastic process X is a modification of Y, and $\mathbb{T} = [0, \infty)$. If both X and Y are almost surely right-continuous, then X and Y are indistinguishable.*

Proof. Let N_X be the set of measure zero where X is not right-continuous, and let N_Y be the corresponding set for Y. Since X is a modification of Y, for all $r \in \mathbb{Q}_+$ we have $P(\{X_r(\omega) \neq Y_r(\omega)\}) = 0$. This implies that $N := \cup_{r \in \mathbb{Q}_+} \{X_r(\omega) \neq Y_r(\omega)\}$ is a countable union of sets of measure zero, and by the σ-additivity of the probability measure, $P(N) = 0$. Now, for any $t \in \mathbb{T}$, let $\{t_n\} \subset \mathbb{Q}_+$ be a decreasing sequence such that $t_n \to t$, as $n \to \infty$. By construction we have that

$$X_{t_n}(\omega) = Y_{t_n}(\omega), \quad \forall \omega \notin N \cup N_X \cup N_Y.$$

Then, by the right-continuity of X and Y we have:

$$X_t(\omega) = \lim_{n \to \infty} X_{t_n}(\omega) = \lim_{n \to \infty} Y_{t_n}(\omega) = Y_t(\omega),$$

for all $\omega \notin N \cup N_X \cup N_Y$. The set $N \cup N_X \cup N_Y$ does not depend on t, and hence X and Y are indistinguishable, as claimed. $\square$

Definition 2.5. Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $\mathbb{T} \subset \mathbb{R}_+$. A *filtration* $(\mathcal{F}_t)_{t \in \mathbb{T}}$ of $\mathcal{F}$ is an increasing family of σ-subalgebras of $\mathcal{F}$ so that if $s \leq t$, then $\mathcal{F}_s \subseteq \mathcal{F}_t$. A stochastic process X is said to be *adapted* to $(\mathcal{F}_t)_{t \in \mathbb{T}}$ if X_t is $\mathcal{F}_t$-measurable for each $t \in \mathbb{T}$.

The quadruple $(\Omega, \mathcal{F}, (\mathcal{F})_{t \in \mathbb{T}}, P)$ is called a *filtered space.* Note that for a given stochastic process X, setting $\mathcal{F}_t^X := \sigma\{X_s : s \in [0, t]\}$ (the σ-algebra generated by the random variables X_s) for any $t \in D$ yields a filtration on $\mathcal{F}$ and X is trivially adapted to $\mathcal{F}_t^X$. For this reason $\mathcal{F}_t^X$ is called the *natural filtration of* X.

Definition 2.6. Given a filtered space we define for $t \geq 0$

$$\mathcal{F}_{t+} = \cap_{u>t}\mathcal{F}_u$$

and

$$\mathcal{F}_{t-} = \sigma(\cup_{u<t}\mathcal{F}_t),$$

where by $\sigma(\cup_{u<t}\mathcal{F}_t)$ we denote the σ-algebra generated by the collection $\cup_{u<t}\mathcal{F}_t$, in other words, the smallest σ-algebra of $\mathcal{F}$ containing $\cup_{u<t}\mathcal{F}_t$. A filtration is called *right-continuous* if $\mathcal{F}_{t+} = \mathcal{F}_t$ for all $t \in \mathbb{T}$. Similarly a filtration is *left-continuous* if $\mathcal{F}_{t-} = \mathcal{F}_t$ for all $t \in \mathbb{T}$.

Given a filtered space, it is always possible to equip it with a right-continuous filtration, since for any filtration $\mathcal{F}_t$, $\mathcal{F}_{t+}$ is always right-continuous. Indeed, if we set $\mathcal{G}_t := \mathcal{F}_{t+}$ we have

$$\mathcal{G}_{t+} = \cap_{u>t}\mathcal{G}_u = \cap_{u>t}\mathcal{F}_{u+} = \cap_{u>t}\cap_{s>u}\mathcal{F}_s = \cap_{u>s}\mathcal{F}_t = \mathcal{G}_t,$$

so $\mathcal{F}_{t+}$ is right-continuous.

Definition 2.7. A filtered complete probability space $(\Omega, \mathcal{F}, (\mathcal{F})_{t\in\mathbb{T}}, P)$ is said to satisfy the *usual conditions* if

- $\mathcal{F}_0$ contains all the P-null sets of $\mathcal{F}$,
- the filtration $(\mathcal{F})_{t\in\mathbb{T}}$ is right-continuous.

In the sequel *we will always assume that the usual conditions hold,* although often we will state this explicitly for clarity. In our case with the Brownian motion this is a technical requirement that can always be satisfied, for details see [17].

2.2 Conditional expectation

Let $(\Omega, \mathcal{F}, P)$ be a probability space. A random variable $X : \Omega \to \mathbb{R}$ is in $L^p(\Omega, \mathcal{F})$, $p \in [1, \infty)$ denoted by $L^p(P)$, if the mathematical expectation of its absolute value to the pth power

$$\|X\|_p^p = \mathbb{E}|X|^p := \int_\Omega |X|^p dP = \int_\Omega |X(\omega)|^p P(d\omega)$$

is finite. In addition, for $1 \leq p < \infty$, $L^p(P)$ is a Banach space. In particular, $L^2(P)$ is a Hilbert space. Furthermore, since $P(\Omega) = 1$, we always have $L^p(P) \subseteq L^q(P)$ for $q \leq p$.

Remark 2.8. In order to improve readability we will follow the probability theoretic convention of truncating the ω-dependence in integrals and elsewhere when it is clear from the context.

Definition 2.9. Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $X \in L^1(\Omega, \mathcal{F}, P)$. Suppose that $\mathcal{G}$ is σ-subalgebra of $\mathcal{F}$. The *conditional expectation of* X *given* $\mathcal{G}$ is the almost surely unique random variable Y that is $\mathcal{G}$-measurable and satisfies

$$\int_A X dP = \int_A Y dP, \quad \text{for all } A \in \mathcal{G}.$$

We denote Y by $\mathbb{E}_{\mathcal{G}} X$.

Remark 2.10. The fact that $\mathbb{E}_{\mathcal{G}} X$ is well-defined (as well as its uniqueness up to P-measure 1) is a consequence of the Radon-Nikodym theorem. For details one may consult any measure-theoretic real analysis textbook, such as [42].

Remark 2.11. Particularly when considering the space $L^2(\Omega, \mathcal{F}, P)$, the conditional expectation is a projection onto $L^2(\Omega, \mathcal{G}, P)$, when $\mathcal{G} \subset \mathcal{F}$, and the notation chosen for $\mathbb{E}_{\mathcal{G}} X$ is also suggestive of this fact. We sketch here the underlying idea only briefly. Let $X \in L^2(P)$. Note that the integral condition in Definition 2.9 reads, for $A \in \mathcal{G}$,

$$E \mathbb{1}_A X = \mathbb{E} \mathbb{1}_A Y,$$

which using the monotone convergence theorem can be extended from indicators via simple functions to hold for all $\mathcal{G}$-measurable functions Z, so it follows that the above condition is equivalent to:

$$\mathbb{E} XZ = \mathbb{E} YZ, \quad \text{for all } \mathcal{G}\text{-measurable functions } Z.$$

We can rewrite the above equation with the inner product of $L^2(P)$ in the form

$$\mathbb{E} Z(X - Y) = \langle Z, X - Y \rangle_{L^2} = 0,$$

which is equivalent to stating that $X - \mathbb{E}_{\mathcal{G}} X$ is orthogonal to each $\mathcal{G}$-measurable function Z, justifying the projection claim. For a rigorous proof, see [6].

The conditional expectation shares many features with regular mathematical expectation, and first we establish its linearity:

Lemma 2.12. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, and suppose $X, Y \in L^1(\Omega, P)$, and $\alpha, \beta \in \mathbb{R}$. Then for any $\mathcal{G} \subset \mathcal{F}$ we have*

$$\mathbb{E}_{\mathcal{G}}(\alpha X + \beta Y) = \alpha \mathbb{E}_{\mathcal{G}} X + \beta \mathbb{E}_{\mathcal{G}} Y, \quad a.s.$$

Proof. Fix $A \in \mathcal{G}$. Now by definition (2.9):

$$\begin{aligned}\int_A \mathbb{E}_{\mathcal{G}}(\alpha X + \beta Y) dP &= \int_A (\alpha X + \beta Y) dP \\ &= \int_A (\alpha X) dP + \int_A (\beta Y) dP \\ &= \alpha \int_A X dP + \beta \int_A Y dP \\ &= \alpha \mathbb{E}_{\mathcal{G}} X + \beta \mathbb{E}_{\mathcal{G}} Y,\end{aligned}$$

where the second and the third equalities follow from elementary properties of the integral. Since this holds for any $A \in \mathcal{G}$, the proof is finished. □

In fact, conditional expectation has several properties that mirror those of the mathematical one, and are also proved similarly by using well-known properties of mathematical expectation. We list a few of these in the following proposition, in addition to some properties that are specific to conditional expectation:

Proposition 2.13. *(Properties of conditional expectation). Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $X \in L^1(\Omega, \mathcal{F}, P)$. Suppose $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ and $Y \in L^1(\Omega, \mathcal{F}, P)$. Then the following hold almost surely:*

(i) (Monotonicity). If $X \leq Y$ almost surely, then $\mathbb{E}_{\mathcal{G}} X \leq \mathbb{E}_{\mathcal{G}} Y$.

(ii) (Monotone / Dominated convergence). Suppose $\{X_n\}_{n \in \mathbb{N}} \subset L^1(\Omega, \mathcal{F}, P)$ converges almost surely to X, and that either 1) X_n is nondecreasing, or 2) there exists $Y \in L^1(\Omega, \mathcal{F}, P)$ for which $|X_n| \leq Y$ almost surely for every $n \in \mathbb{N}$. Then

$$\mathbb{E}_{\mathcal{G}} X_n \to \mathbb{E}_{\mathcal{G}} X \quad \text{as } n \to \infty.$$

(iii) (Fatou). Suppose $\{X_n\}_{n \in \mathbb{N}} \subset L^1(\Omega, \mathcal{F}, P)$ is a nonnegative sequence. Then

$$\mathbb{E}_{\mathcal{G}} \liminf_{n \to \infty} X_n \leq \liminf_{n \to \infty} \mathbb{E}_{\mathcal{G}} X_n.$$

(iv) (Jensen). Suppose $f : \mathbb{R} \to \mathbb{R}$ is convex, and $f \circ X$ is integrable. Then

$$f(\mathbb{E}_{\mathcal{G}} X) \leq \mathbb{E}_{\mathcal{G}} f \circ X.$$

(v) (Expectation by conditioning). $\mathbb{E}(\mathbb{E}_{\mathcal{G}} X) = \mathbb{E} X$.

(vi) If X is $\mathcal{G}$-measurable, then $\mathbb{E}_{\mathcal{G}} X = X$.

(vii) If X and $\mathcal{G}$ are independent, that is, the sigma-algebras $\sigma(X)$ and $\mathcal{G}$ are independent, then $\mathbb{E}_{\mathcal{G}}X = \mathbb{E}X$.

(viii) If Y is $\mathcal{G}$-measurable, and $XY \in L^1(\Omega, \mathcal{F}, P)$, then $\mathbb{E}_{\mathcal{G}}XY = Y\mathbb{E}_{\mathcal{G}}X, a.s.$

(ix) (Tower property). $\mathbb{E}_{\mathcal{H}}X = \mathbb{E}_{\mathcal{G}}(\mathbb{E}_{\mathcal{H}}X) = \mathbb{E}_{\mathcal{H}}(\mathbb{E}_{\mathcal{G}}X)$.

Proof. As noted above, the properties (i)-(iv) follow rather directly from the proofs of these statements for mathematical expectation, and therefore are omitted.

We establish (v) immediately: simply take $A = \Omega$ in Definition 2.9. For (vi), if X is $\mathcal{G}$-measurable, then by the almost sure uniqueness of the conditional expectation noted in Remark 2.10, $X = \mathbb{E}_{\mathcal{G}}X$.

For (vii), let $A \in \mathcal{G}$. Since $\mathbb{E}X$ is a constant, it is $\mathcal{G}$-measurable. By assumption X and $\mathbb{1}_A$ are independent. Now

$$\int_A \mathbb{E}X dP = \mathbb{E}(\mathbb{E}X\mathbb{1}_A) = \mathbb{E}X\mathbb{E}(\mathbb{1}_A) = \mathbb{E}(X\mathbb{1}_A) = \int_A X dP,$$

where the penultimate equality follows from the fact that for independent integrable random variables X, Y such that XY is integrable, we have $\mathbb{E}(XY) = \mathbb{E}X\mathbb{E}Y$. Thus we have shown that $\mathbb{E}X = \mathbb{E}_{\mathcal{G}}X$.

To prove (viii), first assume that Y is a simple function, and thus has the representation $Y = \sum_{i=1}^n a_i \mathbb{1}_{A_i}$, where $A_i \in \mathcal{G}$. Let $A \in \mathcal{G}$. By the definition of the conditional expectation, and then using linearity of the integral we have:

$$\begin{aligned}\int_A \mathbb{E}_{\mathcal{G}}XY dP &= \int_A XY dP = \int_A X \sum_{i=1}^n a_i \mathbb{1}_{A_i} dP = \sum_{i=1}^n a_i \int_{A \cap A_i} X dP \\ &= \sum_{i=1}^n a_i \int_{A\cap A_i} \mathbb{E}_{\mathcal{G}}X dP = \int_A \sum_{i=1}^n a_i \mathbb{1}_{A_i} \mathbb{E}_{\mathcal{G}}X dP = \int_A Y\mathbb{E}_{\mathcal{G}}X dP,\end{aligned}$$

and since every measurable function is a limit of non-negative simple functions, applying (ii) yields the claim.

To prove (ix) note first that by definition $\mathbb{E}_{\mathcal{H}}X$ is $\mathcal{H}$-measurable, and thus also $\mathcal{G}$-measurable, since $\mathcal{H} \subset \mathcal{G}$. Therefore by (vi) proved above, $\mathbb{E}_{\mathcal{G}}(\mathbb{E}_{\mathcal{H}}X) = \mathbb{E}_{\mathcal{H}}X$, which proves the first equality. To prove the second, we show that $\mathbb{E}_{\mathcal{H}}(\mathbb{E}_{\mathcal{G}}X) = \mathbb{E}_{\mathcal{H}}X$. By the definition of $\mathbb{E}_{\mathcal{H}}X$, for any $A \in \mathcal{H}$ we have

$$\int_A X dP = \int_A \mathbb{E}_{\mathcal{H}}X dP.$$

Next, by the definition of $\mathbb{E}_{\mathcal{G}}X$ for any $B \in \mathcal{G}$ we have

$$\int_B XdP = \int_B \mathbb{E}_{\mathcal{G}}XdP,$$

but since $\mathcal{H} \subset \mathcal{G}$ by assumption, the above equality holds a fortiori for each $B \in \mathcal{H}$, which proves the claim.

□

2.3 Martingales and the Doob inequalities

Definition 2.14. Let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $(\mathcal{F}_t)_{t\in\mathbb{T}}$. A real-valued stochastic process $X = \{X_t\}_{t\in\mathbb{T}}$ is called a *submartingale* (with respect to $(\mathcal{F}_t)_{t\in\mathbb{T}}$ if it is adapted to $(\mathcal{F}_t)_{t\in\mathbb{T}}$, $X_t \in L^1(\Omega, P)$, for all $t \in \mathbb{T}$, and $\mathbb{E}_{\mathcal{F}_s}(X_t - X_s) \geq 0$ whenever $s \leq t$. X is called a *supermartingale* if $-X$ is a submartingale, and X is called a *martingale* if X is both a submartingale and a supermartingale.

Note that in the above definition the parentheses are not strictly necessary, but are there to mainly aid intuition to think of the expectation of increments given the information already available. In fact, since by assumption X_s is F_s-measurable, so by (vi) in Proposition 2.13 we have $\mathbb{E}_{\mathcal{F}_s}X_s = X_s$ a.s., so for a submartingale the conditional expectation condition in the definition becomes $\mathbb{E}_{\mathcal{F}_s}X_t \geq X_s$.
We will use the following basic result about martingales extensively throughout, sometimes without specific mention:

Lemma 2.15. *Let X be a martingale with respect to a filtration $(\mathcal{F}_t)_{t\in\mathbb{T}}$, and let $f : \mathbb{R} \to \mathbb{R}$ be a convex function such that $\mathbb{E}f(X)_t < \infty$ for every $t \in \mathbb{T}$. Then $f(X)_t$ is a submartingale.*

Proof. We show first that $f(X)_t$ is adapted to $\mathcal{F}_t$ for all $t \in \mathbb{T}$. Note that since X is a martingale, we have that the mapping $\omega \mapsto X_t(\omega)$ is $\mathcal{F}_t$-measurable for all $t \in \mathbb{T}$ by definition. Also it is well-known that a real-valued convex function on $\mathbb{R}$ is continuous on its domain. Hence f is continuous, and therefore a measurable function on $\mathbb{R}$, so for any open set $G \subset \mathbb{R}$, we have $f^{-1}G \in \mathcal{B}(\mathbb{R})$, where $\mathcal{B}(\mathbb{R})$ denotes the Borel-sigma algebra on $\mathbb{R}$. Therefore using this fact, we have, for any $t \in \mathbb{T}$ that

$$f((X)_t)^{-1}G = X_t^{-1}f^{-1}G \in \mathcal{F}_t,$$

since X_t is adapted to $\mathcal{F}_t$. The above equation shows that for any $t \in \mathbb{T}$, $f(X)_t$ is adapted to $\mathcal{F}_t$.
Next, as by our assumption $f(X)_t \in L^1(P)$ for all $t \in \mathbb{T}$, what remains is to show that if $s \leq t$, we have

$$\mathbb{E}_{\mathcal{F}_s}f(X)_t \geq f(X)_s.$$

Note that since X is a martingale by assumption, we have that when $s \leq t$, the conditional expectation condition is satisfied:

$$X_s = \mathbb{E}_{\mathcal{F}_s} X_t.$$

Now applying f to the above equality and then using (iv) (Jensen) of Proposition 2.13 we have:

$$f(X)_s = f(\mathbb{E}_{\mathcal{F}_s} X_t) \leq \mathbb{E}_{\mathcal{F}_s} f(X)_t,$$

which is precisely the conditional expectation condition for submartingales as in Definition 2.14. Thus we have shown that $f(X)_t$ is a submartingale with respect to $\mathcal{F}_t$, and the proof is complete. □

Definition 2.16. Let $(\Omega, \mathcal{F}, P)$ be a measure space, and $(\mathcal{F}_t)_{t\in\mathbb{T}}$ its filtration. A random variable $T : \Omega \to \overline{\mathbb{R}}_+$ is called a *stopping time* if

$$\{T \leq t\} := \{\omega \in \Omega : T(\omega) \leq t\} \in \mathcal{F}_t \quad \text{for every } t \in \mathbb{T}.$$

Note that any real constant λ is a stopping time.

Definition 2.17. Let $(\Omega, \mathcal{F}, P)$ be a probability space, let X be a stochastic process, and T a stopping time with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{T}}$ of $\mathcal{F}$. We define the stopped process X_t^T for any $t \geq 0$ by

$$X_t^T(\omega) := X_{t\wedge T(\omega)}(\omega).$$

Next we state these following two fundamental lemmas of martingale theory without proof. For details, see for instance [41].

Lemma 2.18. *(Doob submartingale inequality, discrete case). Suppose $\{X_n\}_{n\in\mathbb{N}}$ is a submartingale on a filtered space $(\Omega, \mathcal{F}, (\mathcal{F})_{t\in\mathbb{N}}, \mathbb{P})$. Then for every $\lambda > 0$,*

$$\lambda P(\sup_n X_n \geq \lambda) \leq \sup_n \mathbb{E} X_n^+,$$

where X^+ denotes the positive part of X.

Lemma 2.19. *(Doob L^p Inequality, discrete case). Suppose X is a martingale or a nonnegative submartingale. Then, for $p \in (1, \infty]$, we have*

$$\sup_n |X_n| \in L^p \text{ if and only if } \sup_n \|X_n\|_p < \infty.$$

Furthermore, for $p \in (1, \infty)$ and p, q Hölder conjugates, we have the inequality

$$\|\sup_n |X_n|\|_p \leq q \sup_n \|X_n\|_p.$$

Lemma 2.20. *(Doob submartingale inequality, continuous case). Let X be a submartingale on a filtered space $(\Omega, \mathcal{F}, (\mathcal{F})_{t\in\mathbb{T}}, \mathbb{P})$. Let S be a countable dense subset of $\mathbb{R}_+$, and $I = [a, b) \subset \mathbb{R}_+$. Then for any $\lambda > 0$,*

$$\lambda P(\sup_{t\in I\cap S} X_t > \lambda) \leq \mathbb{E}X_b^+.$$

In particular if X is a right-continuous process, then $I \cap S$ can be replaced with I.

Proof. Since S is a countable dense subset of $\mathbb{R}_+$, consider $\{s_n\}_{n\in\mathbb{N}}$, the arbitrary renumbering of elements of $I \cap S$. For any fixed $N \in \mathbb{N}$, let $\mathcal{T} := \{t_0, t_1, ..., t_N\}$ be a increasing rearrangement of the collection $\{s_0, s_1, ..., s_N\}$. Now $(\mathcal{F}_t)_{t\in\mathcal{T}}$ is a filtration, and furthermore $\{X_t\}_{t\in\mathcal{T}}$ is a submartingale with respect to it. Clearly

$$\max_{t\in\{s_0,...,s_N\}} X_t \nearrow \sup_{t\in I\cap S} X_t, \quad \text{as } N \to \infty.$$

Also for any $\alpha > 0$ we have

$$P(\max_{t\in\{s_0,...,s_N\}} X_t > \alpha\}) = P(\max_{t\in\{t_0,...,t_N\}} X_t > \alpha\}).$$

Hence by the continuity of the probability measure P, and the discrete Doob submartingale inequality in Lemma 2.18, we have, for any $\lambda > 0$:

$$\begin{aligned}
\lambda P(\{\sup_{t\in I\cap S} X_t > \lambda\}) &= \lambda \lim_{N\to\infty} P(\{\max_{t\in\{t_0,...,t_N\}} X_t > \lambda\}) \\
&\leq \lim_{N\to\infty} \mathbb{E}(X_{t_N}^+) \\
&= \mathbb{E}(X_b^+),
\end{aligned}$$

where the last equality follows from the monotone convergence theorem. This proves the inequality in the claim.
Now, if X is right-continuous, then $\sup_{t\in I} X_t(\omega) = \sup_{t\in I\cap S} X_t(\omega)$, so $I \cap S$ can be replaced with I in the above chain of inequalities, and this finishes the proof. □

Lemma 2.21. *(Doob's L^p Inequality, continuous case). Let X be a right-continuous martingale on $[0, \infty) \times \Omega$). Then for every $t > 0$ and $p \in (1, \infty)$ we have the inequality*

$$\left\| \sup_{0\leq s\leq t} X_s \right\|_{L^p(\Omega)} \leq q \sup_{0\leq s\leq t} \left\| X_s \right\|_{L^p(\Omega)},$$

where p, q are Hölder conjugates.

Proof. This is proved similarly as Lemma 2.20 above, by using the discrete version of the claim (Lemma 2.19). We refer the interested reader to [41] for the details. □

2.4 Brownian motion

The purpose of this section is to give the definition of a mathematical object called Brownian motion and prove some of its simple properties needed in the following chapters. There are various approaches to proving its existence, and here we simply point the interested reader to [21] for details.

Definition 2.22. Let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $(\mathcal{F}_t)_{t\in\mathbb{T}}$. A stochastic process $\{W_t\}_{t\in\mathbb{T}}$ is a (one-dimensional) *Brownian motion* if it satisfies

(i) $W_0 = 0$ almost surely,

(ii) For all $0 \leq s < t$, the random variable $W_t - W_s$ is normally distributed with mean 0 and variance $t - s$, i.e. $W_t - W_s \sim N(0, t-s)$,

(iii) $W_t(\omega)$ has independent increments, so that for any $u \geq 0$, the random variable $W_{t+u} - W_t$ is independent of $\sigma(W_s : s \leq t)$,

(iv) Almost all sample paths of $W_t(\omega)$ are continuous functions.

Proposition 2.23. *(Some simple properties of Brownian motion). Let W_t be a fixed Brownian motion, and $t \in \mathbb{T}$. Then*

(i) For any $t \geq 0$, the random variable W_t is normally distributed with mean 0 and variance t.

(ii) (Covariance) For $s, t \geq 0$, we have $Cov(W_tW_s) = \mathbb{E}W_tW_s = t \wedge s$.

(iii) (Translation invariance). For a fixed $a \geq 0$, the stochastic process $\tilde{W}_t := W_{t+a} - W_a$ is also a Brownian motion.

(iv) (Scaling invariance). For any real number $\lambda > 0$, the stochastic process $\tilde{W}_t = \frac{W_{\lambda t}}{\sqrt{\lambda}}$ is also a Brownian motion.

Proof. We note that (i) follows immediately from Definition 2.22, by choosing $s = 0$ in (ii), and then noting that by (i) $W_0 = 0$ almost surely, so $W_t = W_t - W_0 \sim N(0,t)$ by (ii) again.
To prove (ii), note that for $s \leq t$ we can write

$$\mathbb{E}W_sW_t = \mathbb{E}\Big[W_s(W_t - W_s) + W_s^2\Big],$$

and by (iii) of Definition 2.22 the increment $W_t - W_s$ is independent of W_s, so it follows that $\mathbb{E}(W_s(W_t - W_s)) = \mathbb{E}W_s\mathbb{E}(W_t - W_s)$, thus by (ii) and (iii) in Definition 2.22 we have

$$\mathbb{E}W_sW_t = \mathbb{E}\left[W_s(W_t - W_s) + W_s^2\right] = \mathbb{E}W_s\mathbb{E}(W_t - W_s) + \mathbb{E}W_s^2 = 0 + s,$$

hence by symmetry, for $s, t \geq 0$, $\mathbb{E}W_sW_t = t \wedge s$.
To prove (iii), we verify properties (i)-(iv) of Definition 2.22 for $\tilde{W}_t := W_{t+a} - W_a$. Since W_t is a Brownian motion, we have immediately that (i) and (iv) in Definition 2.22 are satisfied. Next, we need to show that $\tilde{W}_t - \tilde{W}_s \sim N(0, t-s)$ for any $0 \leq s < t$. Let $s \leq t$, and then $\tilde{W}_t - \tilde{W}_s = W_{t+a} - W_{s+a}$. Hence $\mathbb{E}(\tilde{W}_t - \tilde{W}_s) = 0$, by (ii) of Definition 2.22, and using (ii) of this proposition, we have

$$\begin{aligned}
\operatorname{Var}(\tilde{W}_t - \tilde{W}_s) &= \operatorname{Var}(W_{t+a} - W_{s+a}) = \operatorname{Var} W_{t+a} + \operatorname{Var} W_{s+a} - 2\operatorname{Cov}(W_{t+a}, W_{s+a}) \\
&= t + s - 2s \wedge t \\
&= t - s,
\end{aligned}$$

so $\tilde{W}_t - \tilde{W}_s \sim N(0, t-s)$, and we have verified (ii) of Definition 2.22. What remains is to show (iii) in Definition 2.22. We need to show that for all $u \geq 0$, the random variable $\tilde{W}_{t+u} - \tilde{W}_t = W_{t+a+u} - W_{t+a}$ is independent of $\sigma(\tilde{W}_s : s \leq t)$. It is enough to note that since the Brownian motion W_t has this property for all $t \geq 0$, we can simply choose $t = t + a$, and hence (iii) of Definition 2.22 is satisfied as well, and thus $\tilde{W}_t$ is a Brownian motion.
The proof of (iv) is similar to (iii) above, and is omitted. □

Proposition 2.24. *Let $T \in \mathbb{R}$. The Brownian motion W_t is a martingale with respect to its natural filtration given by $\mathcal{F}_t := \sigma\{W_s : s \leq t\}$ for every $t \in [0, T]$.*

Proof. The adaptability condition is obvious. Note that by (i) in Proposition 2.23, $W_t \sim N(0, t)$, hence $\mathbb{E}W_t^2 = t \leq T < \infty$, so $W_t \in L^2(P)$ for all $t \in [0, T]$, and thus $W_t \in L^1(P)$. Next, we need to show that if $s \leq t$, we have

$$\mathbb{E}_{\mathcal{F}_s}W_t - W_s = 0.$$

Note that by using the linearity of the conditional expectation we can write

$$\mathbb{E}_{\mathcal{F}_s}W_t = \mathbb{E}_{\mathcal{F}_s}\Big(W_t - W_s + W_s\Big) = \mathbb{E}_{\mathcal{F}_s}\Big(W_t - W_s\Big) + \mathbb{E}_{\mathcal{F}_s}W_s. \tag{2.25}$$

Now since W_s is trivially $\mathcal{F}_s$-measurable, we have by property (vi) of Proposition 2.13 that $\mathbb{E}_{\mathcal{F}_s}W_s = W_s$. Also by (iii) in Definition 2.22, $W_t - W_s$ is independent of $\mathcal{F}_s$, so by (vii) of Proposition 2.13, we have $\mathbb{E}_{\mathcal{F}_s}\Big(W_t - W_s\Big) = \mathbb{E}\Big(W_t - W_s\Big) = 0$, where the last equality is again due to (ii) of Definition 2.22. Now it follows from (2.25) that we have $\mathbb{E}_{\mathcal{F}_s}W_t = W_s$, and the claim is proved. □

2.5 Kolmogorov Continuity Theorem

The aim of this section is to establish the Kolmogorov Continuity Theorem, first proved by Andrey Kolmogorov in 1934. This theorem is also referred to as Kolmogorov-Chentsov Theorem, particularly in a more general setting of metric spaces. Here we will follow an approach taken in [10]. Before stating and proving the claim of this theorem, we will state and prove a technical lemma:

Lemma 2.26. *Let $\mathcal{D}$ be the set of all dyadic rationals in $[0,1]$, that is, the real numbers in $[0,1]$ of the form $\frac{i}{2^n}$ for some integer $n \geq 1$, and some $i \in \{0,1,...,2^n-1\}$. Let $f : \mathcal{D} \to \mathbb{R}$, and assume that there exists $q > 0$ and a constant $K \in \mathbb{R}$, such that for every integer $n \geq 1$ and every $i \in \{1,2,...,2^n-1\}$,*

$$\Big|f\Big(\frac{i-1}{2^n}\Big) - f\Big(\frac{i}{2^n}\Big)\Big| \leq K\frac{1}{2^{nq}}.$$

Then we have, for every $s,t \in \mathcal{D}$, that

$$|f(s)-f(t)| \leq \frac{2K}{1-2^{-q}}|t-s|^q.$$

Proof. Fix $s,t \in \mathcal{D}$, and let $p \geq 1$ be the smallest integer such that $2^{-p} \leq t-s$, and let $k \geq 0$ be the smallest integer such that $k2^{-p} \geq s$. Now, since each dyadic rational has a finite binary expansion, we may write

$$\begin{aligned} s &= k2^{-p} - \epsilon_1 2^{-p-1} - ... - \epsilon_l 2^{-p-l} \\ t &= k2^{-p} + \epsilon'_0 2^{-p} + \epsilon'_1 2^{-p-1} + ... + \epsilon'_m 2^{-p-m}, \end{aligned}$$

where $l,m \in \mathbb{N}$, and $\epsilon_i, \epsilon'_j = 0$ or 1, for every $i \in [1,l]$, and $j \in [0,m]$. Next, we define

$$\begin{aligned} s_i &= k2^{-p} - \epsilon_1 2^{-p-1} - ... - \epsilon_i 2^{-p-i} \\ t_j &= k2^{-p} + \epsilon'_0 2^{-p} + \epsilon'_1 2^{-p-1} + ... + \epsilon'_j 2^{-p-j}, \end{aligned}$$

for every $i \in [0,l]$ and $j \in [0,m]$. Note that $s = s_l, t = t_m$. Hence applying the inequality in the assumption of the lemma to the pairs $(s_0, t_0), (s_{i-1}, s_i), (t_{i-1}, t_i)$, yields:

$$
\begin{aligned}
|f(s) - f(t)| = |f(s_l) - f(t_m)| &\leq |f(s_0) - f(t_0)| \\
&+ \sum_{i=1}^{l} |f(s_{i-1} f(s_i)| + \sum_{j=1}^{m} |f(t_{j-1}) - f(t_j)| \\
&\leq \frac{K}{2^{pq}} + \sum_{i=1}^{l} \frac{K}{2^{(p-i)q}} + \sum_{j=1}^{m} \frac{K}{2^{(p-j)q}} \\
&\leq 2K(1 - 2^{-q})^{-1} 2^{-pq} \\
&\leq 2K(1 - 2^{-q})^{-1} (t - s)^q,
\end{aligned}
$$

with the final inequality following from $2^{-p} \leq t - s$. The proof is complete. □

Theorem 2.27. *(Kolmogorov Continuity Theorem). Let $I \subset \mathbb{R}$ be a bounded interval, and $X = (X_t)_{t \in I}$ be a stochastic process on $(\Omega, \mathcal{F}, P)$ taking values in $\mathbb{R}$. Assume that*

$$\mathbb{E}(|X_t - X_s|^\alpha) \leq c|t - s|^{1+\beta}, \quad \forall s, t \in I$$

holds for some constant $c > 0$ and $\alpha, \beta > 0$. Then X has a modification, denoted by $\tilde{X}$, that has Hölder continuous sample paths on I for every $q \in (0, \frac{\beta}{\alpha})$.

Proof. Fix $q \in (0, \frac{\beta}{\alpha})$. Since $I \subset \mathbb{R}$ is a bounded interval, it is sufficient to prove the theorem when $I = [0, 1]$. Now let $s, t \in I$, and then for $\lambda > 0$ the Chebyshev inequality and the assumption imply:

$$
\begin{aligned}
P(|X_s - X_t| \geq \lambda) &\leq \frac{1}{\lambda^\alpha} \mathbb{E}|X_s - X_t|^\alpha \\
&\leq \frac{1}{\lambda^\alpha} c|t - s|^{1+\beta}. \qquad (2.28)
\end{aligned}
$$

Now, fix $n \in \mathbb{N}$, and let $s = \frac{i-1}{2^n}$, and $t = \frac{i}{2^n}$, with $i \in \{1, 2, ..., 2^n\}$. Choose $\lambda = 2^{-nq}$. Note that now we obtain from the estimate (2.28) that

$$P(|X_s - X_t| \geq 2^{-nq}) \leq \frac{1}{2^{(-nq)\alpha}} c \left| \frac{i}{2^n} - \frac{i-1}{2^n} \right|^{1+\beta} = c 2^{nq\alpha} 2^{-n(1+\beta)}.$$

Hence taking the union over i we have

$$P\Big(\bigcup_{i=1}^{2^n} |X_s - X_t| \geq 2^{-nq} \Big) \leq \sum_{i=1}^{2^n} c 2^{nq\alpha} 2^{-n(1+\beta)} = c 2^{n(q\alpha - \beta)}.$$

By assumption $q\alpha - \beta < 0$, and from this it follows that

$$\sum_{n=1}^{\infty} P\Big(\bigcup_{i=1}^{2^n} |X_s - X_t| \geq 2^{-nq} \Big) < \infty.$$

Therefore by the Borel-Cantelli lemma $\bigcup_{i=1}^{2^n} |X_s - X_t| \geq 2^{-nq}$ for only finitely many n, in other words, there exists $n_0 = n_0(\omega)$, such that $|X_s - X_t| \leq 2^{-nq}$, for all $n \geq n_0$, and for all $i \in \{1, 2, ..., 2^n\}$. Now define

$$K_q(\omega) := \sup_{n \geq 1} \Big(\sup_{i \in \{1,2,...,2^n\}} \frac{|X_s - X_t|}{2^{-nq}} \Big).$$

From the above it follows that $K_q(\omega) < \infty$ almost surely. Now consider the set $\mathcal{D}$ as in Lemma 2.26. We have established above that for any $s, t \in \mathcal{D}$, and $f := X$, the assumption of Lemma 2.26 is satisfied. Hence we have for any $s, t \in \mathcal{D}$:

$$|X_s - X_t| \leq C_q(\omega) |t - s|^q,$$

with $C_q(\omega) := \frac{2K_q(\omega)}{1-2^{-q}}$. Hence we have shown that $t \overset{g}{\mapsto} X_t$ is q-Hölder continuous on $\mathcal{D}$, and therefore also uniformly continuous on $\mathcal{D}$. Note that combining this with the facts that $\mathcal{D}$ is dense in $[0, 1]$, and $\mathbb{R}$ is complete, it follows from a simple triangle inequality argument that we can extend g uniquely to $[0, 1]$, and the extension is also q-Hölder continuous on $[0, 1]$. Hence defining

$$\tilde{X}_t(\omega) = \begin{cases} \lim_{s \to t} X_t(\omega), & \text{if } K_q(\omega) < \infty \\ 0, & \text{if } K_q(\omega) = \infty \end{cases}$$

for all $t \in [0, 1]$, we have that $t \mapsto \tilde{X}_t$ is Hölder continuous on I.
What remains is to show that $\tilde{X}_t$ is indeed a modification of X_t. By definition we need to show that for all $t \in I$, $\tilde{X}_t = X_t$ almost surely. By the assumption in the theorem we have that $X_s \to X_t$ in $L^1(P)$ as $s \to t$, and hence this convergence holds in probability as well. But this is precisely how we have defined $\tilde{X}_t$, so it must hold that $\tilde{X}_t$ is a modification of X_t, and the proof is finished. □

As a typical corollary of the preceding theorem we establish an important continuity property of the Brownian motion. By definition the sample paths of W are almost surely continuous, but in fact something more is true:

Corollary 2.29. *Let $I \subset \mathbb{R}$ a bounded interval. The Brownian motion W has q-Hölder continuous sample paths on I almost surely, for any $q \in (0, \frac{1}{2})$.*

Proof. Note that for $s \leq t$, the random variable $W_t - W_s \sim N(0, t-s)$ by definition, so for any $k \in \mathbb{N}$ the computation of the Gaussian integral yields:

$$\mathbb{E}W_t^{2k+1} = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t}} x^{2k+1} e^{-\frac{x^2}{2t}} dx = 0,$$

and

$$\mathbb{E}W_t^{2k} = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t}} x^{2k} e^{-\frac{x^2}{2t}} dx = t^k \frac{2^k \Gamma(k+1/2)}{\sqrt{\pi}}.$$

Hence we have:

$$\mathbb{E}|W_t - W_s|^{2k} = \frac{\Gamma(k+1/2)}{\sqrt{\pi}} |t-s|^k,$$

so we apply Theorem 2.27, with $\alpha = 2k, 1+\beta = k$, and obtain that for $k \in \mathbb{N}$, W_t has a modification with Hölder continuous sample paths for any $q \in (0, \frac{k-1}{2k})$. Furthermore, since $\frac{k-1}{2k} \to \frac{1}{2}$, as $k \to \infty$, the result holds for any $q \in (0, \frac{1}{2})$, which was to be proved. □

Remark 2.30. The previous theorem can be extended for Gaussian processes. A stochastic process $\{X_t\}_{t \in \mathbb{T}}$ is called Gaussian, if for any finite subset $\{t_i\}_{i=1}^n \subset \mathbb{T}$, and any $\{a_i\}_{i=1}^n \subset \mathbb{R}$, all linear combinations of the form $a_i X_{t_i}$ are normally distributed. Hence Brownian motion is a Gaussian process. We have the following result, close in spirit to Theorem 2.27, and yielding also its converse for Gaussian processes:

Theorem 2.31. *A Gaussian process $X = \{X_t\}_{t \in \mathbb{T}}$ is Hölder continuous of any order $a < H$ if and only if there exist constants C_ϵ such that for all $s, t \in \mathbb{T}$, we have:*

$$\sqrt{\mathbb{E}|X_t - X_s|^2} \leq C_\epsilon |t-s|^{H-\epsilon}, \quad \text{for all } \epsilon > 0.$$

For proof, see [1].

Chapter 3

Stochastic Integrals with respect to Brownian motion

3.1 Itô isometry for simple left-continuous processes

In this chapter we construct the stochastic integral with respect to a Brownian motion for a suitable class of stochastic processes. To motivate the nature of the construction, we first make a small digression; put more bluntly, we try to make it clear why a more elementary approach will not work in our case.

In basic real analysis one learns that if a function $g : [a, b] \to \mathbb{R}$ is of bounded variation and (for instance) left-continuous, then for a continuous function $f : [a, b] \to \mathbb{R}$ we may define the Riemann-Stieltjes integral as a limit of the sum

$$\sum_{i=1}^{n} f(c_i)\Big(g(t_i) - g(t_{i-1})\Big),$$

as the length of the partition

$$a = t_0 < t_1 < \cdots < t_n = b$$

of $[a, b]$ tends to zero, with c_i an arbitrary point in each subinterval $[t_{i-1}, t_i]$. This integral is typically denoted by $\int_a^b f(x)dg(x)$, and particularly in the more general setting of Lebesgue-Stieltjes integration with $f \in L^1([a, b])$, we may think of the function g inducing a measure on $[a, b]$. In fact, it turns out to be that the only class of functions g, for which an integral of this kind can be defined, are precisely of bounded variation (for proof, see [32]).

It is worth pointing out that knowledge of this last fact implies the failure of Riemann-Stieltjes (or Lebesgue-Stieltjes) integration with respect to the Brownian motion. This follows since Brownian motion is not of bounded variation on any bounded set of $\mathbb{R}$. Indeed, for a partition π_n of an interval $[a, b] \subset \mathbb{R}$, for which $|\pi_n| \to 0$, it can be shown (see, for instance [21] for proof) that the quadratic variation of the Brownian motion,

$$\sum_{t_i^n \in \pi_n} (W_{t_i^n} - W_{t_{i-1}^n})^2$$

tends to $b - a$ in $L^2(P)$. Hence the integral with respect to the random measure induced by the Brownian motion cannot be defined for every $\omega \in \Omega$.
With these considerations in mind, throughout this section we fix a probability space $(\Omega, \mathcal{F}, P)$ with a Brownian motion W, satisfying the usual conditions (UC), restated here for convenience:

- The probability space $(\Omega, \mathcal{F}, P)$ is complete.
- The filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ is right-continuous,
- The filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ is augmented, that is, it contains all the null sets in $(\Omega, \mathcal{F}, P)$,
- The Brownian motion W is adapted to the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$.

It is vital to note that although the above conditions are technical requirements imposed in order to achieve the mathematical construction we desire, they are not vacuous. Indeed, see for instance [17] for a constructive proof of how to obtain such a probability space.

Definition 3.1. Let $L_0(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, P) := L_0$ be the collection of all *bounded adapted left-continuous simple processes* on the corresponding filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, P)$. That is, there exists a nonnegative strictly increasing sequence of real numbers $\{t_k : k \in \mathbb{N}\}$ with $t_0 = 0$, and $\lim_{k\to\infty} t_k = \infty$ and a sequence of random variables $\{\xi_k\}_{k\in\mathbb{N}}$ such that $|\xi_k(\omega)| \leq K$ for every $\omega \in \Omega$ and for some $K \in [0, \infty)$, and ξ_i is $\mathcal{F}_{t_i}$-measurable for every $i \in \mathbb{N}$, and $X \in L_0$ has the form

$$X_t(\omega) = \xi_0(\omega)\mathbb{1}_{\{0\}}(t) + \sum_{i=0}^{\infty} \xi_i(\omega)\mathbb{1}_{(t_i, t_{i+1}]}(t).$$

Definition 3.2. For all $X \in L_0$, we define the *Itô integral of X with respect to W* to be the process given by

$$I_t(X) = \int_0^t X_s dW_s := \sum_{i=1}^{n} \xi_i(\omega)(W_{t_{i+1}\wedge t} - W_{t_i\wedge t})(\omega), \quad t \geq 0.$$

Furthermore we define

$$\int_a^b X_s dW_s := \int_0^b X_s dW_s - \int_0^a X_s dW_s, \quad a, b \geq 0.$$

To verify that the integral in Definition 3.2 is indeed well-defined we must show that it does not depend on the particular representation chosen for $X \in L_0$. So, let

$$X_t(\omega) = \xi_0(\omega)\mathbb{1}_{\{0\}}(t) + \sum_{i=0}^{\infty} \xi_i(\omega)\mathbb{1}_{(t_i,t_{i+1}]}(t),$$

and also

$$X_t(\omega) = \zeta_0(\omega)\mathbb{1}_{\{0\}}(t) + \sum_{j=0}^{\infty} \zeta_j(\omega)\mathbb{1}_{(t_j,t_{j+1}]}(t),$$

with ξ_i being $\mathcal{F}_{t_i}$-measurable for every $i \in \mathbb{N}$, and for ζ_j likewise. Note that we have $(a, b] \cap [c, d] = (a \vee c, b \wedge d]$ for all $a, b, c, d \in \mathbb{R}$. Hence we have

$$X_t(\omega) = \xi_0(\omega)\mathbb{1}_{\{0\}}(t) + \sum_{i=0,j=0}^{\infty} \xi_i(\omega)\mathbb{1}_{(t_i,t_{i+1}]}(t)\mathbb{1}_{(t_j,t_{j+1}]}(t),$$

and also

$$X_t(\omega) = \zeta_0(\omega)\mathbb{1}_{\{0\}}(t) + \sum_{j=0,i=0}^{\infty} \zeta_j(\omega)\mathbb{1}_{(t_j,t_{j+1}]}(t)\mathbb{1}_{(t_i,t_{i+1}]}(t).$$

Now, without loss of generality we can assume that $(t_i, t_{i+1}] \subset (t_j, t_{j+1}]$ for all $i, j \in \mathbb{N}$. Since $\xi_i = \zeta_j$ whenever $(t_i, t_{i+1}] \cap (t_j, t_{j+1}] \neq \emptyset$, by Definition 3.2 we have for each ω and all $t \geq 0$:

$$\begin{aligned}
\sum_{i=1}^{n} \xi_i(W_{t_{i+1}\wedge t} - W_{t_i\wedge t}) &= \sum_{i=1,j=1}^{n} \xi_i(W_{(t_{i+1}\wedge t_{j+1})\wedge t} - W_{(t_i\vee t_j)\wedge t}) \\
&= \sum_{j=1,i=1}^{n} \zeta_j(W_{(t_{j+1}\wedge t_{i+1})\wedge t} - W_{(t_j\vee t_i)\wedge t}) \\
&= \sum_{j=1}^{n} \zeta_j(W_{t_{j+1}\wedge t} - W_{t_j\wedge t}),
\end{aligned}$$

so for any $X \in L_0$, the Itô integral in Definition 3.2 is well-defined.

Remark 3.3. Recall that in the Riemann-Stieltjes integral, for each interval of the type $[t_{i-1}, t_i]$, the point c_i where the integrand is evaluated can be chosen freely on the interval in question, with no difference in the resulting integral. This is in contrast to the integral defined above, where the point where the random variable ξ is evaluated is specifically chosen to be the left endpoint of the particular interval. This is not an arbitrary choice, as choosing a different point will result in an altogether different integral, with different properties. We will return to this point at the end of this chapter, where we briefly discuss the Stratonovich integral. Finally, note that since $W_0 = 0$ almost surely, the value of ξ_0 plays no role in the above definition of the Itô integral.

Remark 3.4. It is worth pointing out that for any discrete stochastic processes X and M the process defined by

$$(X \cdot M)_n = \begin{cases} 0, & n = 0, \\ \sum_{k=1}^{n} X_k(M_k - M_{k-1}), & n \geq 1 \end{cases}$$

is called a martingale transform. Hence when we take $M \equiv W$, the Itô integral as defined in Definition 3.2 is in fact a martingale transform. We will return to this point of more general integrators than the Brownian motion W in the next section.

Proposition 3.5. *(Elementary properties of the Itô integral).*

(i) The Itô integral is linear, that is, for any $X, Y \in L_0$ and for all $\alpha, \beta \in \mathbb{R}$ we have

$$I_t(\alpha X + \beta Y) = \alpha I_t(X) + \beta I_t(Y).$$

(ii) $I_0 = 0$ on Ω.

(iii) For any $X \in L_0$ and $t \in \mathbb{R}_+$, the stochastic process given by $I_t(X)$ is adapted to $\mathcal{F}_t$.

Proof. Properties (i) and (ii) are obvious from Definition 3.2. For (iii), we want to show that $I_t(X) = \sum_{i=0}^{n} \xi_i(W_{t_{i+1}\wedge t} - W_{t_i \wedge t})$ is $\mathcal{F}_t$-measurable for all $t \in \mathbb{R}_+$. By definition each ξ_i is $\mathcal{F}_{t_i}$-measurable. By construction, the Brownian motion W_t is adapted to $(\mathcal{F}_t)$. It is a well-known fact from real analysis that measurability is preserved under differences and products, so $I_t(X)$ is adapted to $(\mathcal{F}_t)$, and the proof is finished. □

Lemma 3.6. *(Itô Isometry for simple processes). Let $X \in L_0$. Then we have the Itô isometry*

$$\mathbb{E}(|(I_t(X)|^2) = \mathbb{E}\Big(\int_0^t X_s^2 ds \Big). \tag{3.7}$$

Furthermore, the stochastic integral $I_t(X)$ is a continuous martingale with respect to the filtration $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$, and also $\mathbb{E}|I_t(X)|^2 < \infty$. That is, $I_t(X)$ is square-integrable.

Proof. We first prove that the equation stated in the lemma is valid. Note that this also yields the square-integrability of $I_t(X)$. Let $X \in L_0$ so we have:

$$|I_t(X)|^2 = \sum_{i,j=1}^{n} \xi_{t_i}\xi_{t_j}(W_{t_{i+1}\wedge t} - W_{t_i\wedge t})(W_{t_{j+1}\wedge t} - W_{t_j \wedge t}). \tag{3.8}$$

Since the sum in the above equation is finite and expectation is a linear operator, it suffices to consider the two cases for indices separately. First, let $i \neq j$, and we may assume that $i < j$. Now, using property (v) in Proposition 2.13 and the fact that since by definition ξ_i is adapted to the filtration $(\mathcal{F}_{t_i})$, we can apply (viii) of Proposition 2.13 and compute:

$$
\begin{aligned}
\mathbb{E}\Big(\xi_i\xi_j(W_{t_{i+1}\wedge t} - W_{t_i\wedge t})(W_{t_{j+1}\wedge t} - W_{t_j\wedge t})\Big) &= \mathbb{E}\Big(\mathbb{E}_{\mathcal{F}_{t_j}}\xi_i\xi_j(W_{t_{i+1}\wedge t} - W_{t_i\wedge t})(W_{t_{j+1}\wedge t} - W_{t_j\wedge t})\Big) \\
&= \mathbb{E}\Big(\xi_i\xi_j(W_{t_{i+1}\wedge t} - W_{t_i\wedge t})\mathbb{E}_{\mathcal{F}_{t_j}}(W_{t_{j+1}\wedge t} - W_{t_j\wedge t})\Big) \\
&= 0, \qquad (3.9)
\end{aligned}
$$

since $\mathbb{E}_{\mathcal{F}_{t_j}}(W_{t_{j+1}\wedge t} - W_{t_j\wedge t}) = \mathbb{E}(W_{t_{j+1}\wedge t} - W_{t_j\wedge t}) = 0$, by the properties (ii) and (iii) in Definition 2.22 of Brownian motion.
Next, if $i = j$, a similar computation as above yields

$$
\begin{aligned}
\mathbb{E}\Big(\xi_i^2(W_{t_{i+1}\wedge t} - W_{t_i\wedge t})^2\Big) &= \mathbb{E}\Big(\mathbb{E}_{\mathcal{F}_{t_i}}\xi_i^2(W_{t_{i+1}\wedge t} - W_{t_i\wedge t})^2\Big) \\
&= \mathbb{E}\Big(\xi_i^2\mathbb{E}(W_{t_{i+1}\wedge t} - W_{t_i\wedge t})^2\Big) \\
&= \mathbb{E}\Big(\xi_i^2(t_{i+1} - t_i)\Big) \\
&= (t_{i+1} - t_i)\mathbb{E}(\xi_i^2). \qquad (3.10)
\end{aligned}
$$

It follows from inserting equations (3.9) and (3.10) into (3.8) that we have (3.7), as claimed. In order to show that $I_t(X)$ is continuous, we have, by definition, that

$$I_t(X) = \sum_{i=1}^{n} \xi_i(\omega)(W_{t_{i+1}\wedge t} - W_{t_i\wedge t})(\omega).$$

By definition the Brownian motion W is almost surely a continuous stochastic process, so it immediately follows that $t \mapsto I_t(X)$ is also almost surely a continuous function.
What remains is to prove that the Itô integral is a martingale. Note that by (iii) of Proposition 3.5, $I_t(X)$ is adapted. Also as noted earlier, $I_t(X)$ is square-integrable, so from the Cauchy-Schwartz inequality it follows that $\mathbb{E}|I_t(X)| \leq (\mathbb{E}|I_t(X)|^2)^{\frac{1}{2}} < \infty$. Now let $s \leq t$. We want to show that $\mathbb{E}_{\mathcal{F}_s} I_t(X) = I_s(X)$. Note that $\mathcal{F}_s \subset \mathcal{F}_{s\vee t_i}$. So using

property (ix) of Proposition 2.13 with this fact we have that

$$\begin{aligned}
\mathbb{E}_{\mathcal{F}_s}\Big(\sum_{i=1}^n \xi_i(W_{t_{i+1}\wedge t} - W_{t_i\wedge t})\Big) &= \sum_{i=1}^n \mathbb{E}_{\mathcal{F}_s}\Big(\xi_i(W_{t_{i+1}\wedge t} - W_{t_i\wedge t})\Big) \\
&= \sum_{i=1}^n \mathbb{E}_{\mathcal{F}_s}\Big(\mathbb{E}_{\mathcal{F}_{(t_i\vee s)}}\xi_i(W_{t_{i+1}\wedge t} - W_{t_i\wedge t})\Big) \\
&= \sum_{i=1}^n \mathbb{E}_{\mathcal{F}_s}\Big(\xi_i\mathbb{E}_{\mathcal{F}_{(t_i\vee s)}}(W_{t_{i+1}\wedge t} - W_{t_i\wedge t})\Big) \\
&= \sum_{i=1}^n \mathbb{E}_{\mathcal{F}_s}\Big(\xi_i(W_{t_{i+1}\wedge s} - W_{t_i\wedge s})\Big) \\
&= \sum_{i=1}^n \Big(\xi_i(W_{t_{i+1}\wedge s} - W_{t_i\wedge s})\Big) \\
&= I_s(X),
\end{aligned}$$

with the third equality following from (viii) of Proposition 2.13 due to the measurability condition of ξ_i. Hence we have shown that the Itô integral as defined in Definition 3.2 is a martingale, and the proof is complete. □

3.2 Extension to left-continuous processes

In this section we continue with the assumptions (UC) defined at the start of Section 3.1, and the goal is to define the Itô integral for more general processes than the ones in L_0. More specifically, we will define the Itô integral as the $L^2(P)$-limit for a suitable class of stochastic processes X. For simplicity, we will work on $[0, T] \subset \mathbb{R}$, with $T < \infty$. We begin by defining the space of these processes. First, we need a measurability condition:

Definition 3.11. Let X be a left-continuous, bounded, and adapted process on $[0, T] \times \Omega$. The *predictable $\sigma-$algebra $\mathcal{P}$* is defined as

$$\mathcal{P} := \sigma(X : X \text{ is left-continuous, bounded, and adapted to } (\mathcal{F}_t)_{t\in[0,T]}).$$

When a stochastic process $Y : [0, T] \times \Omega \to \mathbb{R}$ is measurable with respect to the predictable $\sigma-$algebra $\mathcal{P}$, we say that Y is a *predictable process.*

In addition to the measurability condition above, we require an integrability condition with respect to $dP \otimes dt$, namely that the process X on $[0, \infty) \times \Omega$ satisfies

$$\mathbb{E}\int_0^T |X_t(\omega)|^2 dt < \infty \tag{3.12}$$

For convenience we denote $L^2_{ad}([0,T]\times\Omega, dt\otimes dP)$ by L^2_{ad}, the space of stochastic processes $X_t(\omega)$ satisfying both Definition 3.11 and equation (3.12).
In order to define $I_t(X)$ for $X \in L^2_{ad}$, we want to have some suitable approximation results which we can then use to extend the results obtained in Section 3.1. We note that although this fundamental idea is always present in the definition of the Itô integral, dating back to the original paper by Itô himself [14], the details, in particular the generality of the presentation, can differ drastically.
We will here focus on constructing the Itô integral with respect to the Brownian motion only, but in fact some intermediary results are presented in a more general manner. As such, our approach is akin to an average of those presented in [35], [3], and [10]. We return to discuss the point of integrating with respect to more general random processes briefly at the end of this chapter, but for a more general treatment one may consult [32], [7] or [41], for instance.
We begin by we stating and proving a simple lemma:

Lemma 3.13. *Let $\mathcal{C}$ be the collection of processes on $[0,\infty)\times\Omega$ of the form*

$$X_t(\omega) = \sum_{i=1}^{n} \alpha_i(\omega)\mathbb{1}_{(a_i,b_i]}(t), \tag{3.14}$$

where α_i is a bounded, $\mathcal{F}_{a_i}$-measurable random variable for each i. Then, the predictable $\sigma-$algebra $\mathcal{P}$ is generated by $\mathcal{C}$, that is, $\sigma(\mathcal{C}) = \mathcal{P}$.

Proof. The inclusion of $\mathcal{C}\subset\mathcal{P}$ is immediate, since if $X\in\mathcal{C}$, then X is bounded, adapted and left-continuous, and thus predictable. Thus by definition $X\in\mathcal{P}$, and we have $\mathcal{C}\subset\mathcal{P}$. In order to show that $\mathcal{P}\subset\sigma(\mathcal{C})$, let X be left-continuous, bounded and adapted. Let $I_{n,k} = ((k-1)2^{-n}, k2^{-n}]$, for $k,n\in\mathbb{N}$, and define a sequence of left-continuous simple processes $\{X^{(n)}\}_{n\in\mathbb{N}}$ by

$$\begin{cases} X^{(n)}(t,\omega) = X(\frac{k-1}{2^n},\omega), & \text{for } t\in I_{n,k},\ \omega\in\Omega, \\ X^{(n)}(0,\omega) = X(0,\omega), & \text{for } \omega\in\Omega. \end{cases} \tag{3.15}$$

Each $X^{(n)}$ is clearly in $\mathcal{C}$. Note that for each $n\in\mathbb{N}$ there exists a unique $k_n\in\mathbb{N}$ such that $t\in I_{n,k_n}$. Thus by (3.15) we have $X^{(n)}(t,\omega) = X(\frac{k_n-1}{2^n},\omega)$. Since $t\in I_{n,k_n}$, from its definition it follows that we have $\frac{k_n-1}{2^n} < t$, and also that

$$t - \Big(\frac{k_n-1}{2^n}\Big) < \frac{k_n}{2^n} - \Big(\frac{k_n-1}{2^n}\Big) = \frac{1}{2^n},$$

for all $n\in\mathbb{N}$. Thus $\frac{k_n-1}{2^n} \nearrow t$, as $n\to\infty$, and therefore by the left-continuity of X we have that

$$\lim_{n\to\infty} X^{(n)}(t,\omega) = \lim_{n\to\infty} X\Big(\frac{k_n-1}{2^n},\omega\Big) = X(t,\omega).$$

In addition, as by assumption X is $(\mathcal{F}_t)$-adapted, we have that

$$X((k-1)2^{-n}, \omega) \in \mathcal{F}_{(k-1)2^{-n}},$$

for a fixed ω. Hence for any $t \in I_{n,k}$, we have

$$X^{(n)}(t, \omega) \in \mathcal{F}_t,$$

since

$$\mathcal{F}_{(k-1)2^{-n}} \subset \mathcal{F}_t,$$

due to the fact that $(\mathcal{F}_t)_{t\in[0,T]}$ is a filtration. We have shown that we can approximate a member of $\mathcal{P}$ by $X \in \mathcal{C}$. Since measurability is preserved under limits, the σ-algebra generated by $\mathcal{C}$ contains $\mathcal{P}$, and the claim is proved. □

Lemma 3.16. *Let $X \in L^2_{ad}$. Then there exists a sequence of processes $\{X^{(n)}\}_{n\in\mathbb{N}}$ of the form* (3.14) *for which*

$$\lim_{n\to\infty} \mathbb{E} \int_0^T |X_t^{(n)} - X_t|^2 dt = 0.$$

Proof. Setting

$$\|X\|_2 = \left(\mathbb{E} \int_0^T X_t^2 dt \right)^{1/2}$$

for $X \in L^2_{ad}$ we obtain a norm on $L^2([0, T] \times \Omega, dt \otimes dP)$. By Lemma 3.13 we can approximate X with processes of the form given in (3.14). Indeed, define for each $n, k \in \mathbb{N}$ the sequence:

$$X_s^{(n)} = \begin{cases} X_{\frac{k-1}{2^n}}, & (k-1)2^{-n} < s \le k2^{-n} \\ 0, & s \ge n. \end{cases}$$

Now the dominated convergence theorem for L^2-spaces yields the claim. □

Lemma 3.17. *Let $(\Omega, (\mathcal{F}_t)_{t\in[0,T]}, \mathcal{F}, P)$ be a filtered space. Define $\mathcal{M}^{2,c}$ as the vector space of martingales X with respect to $(\mathcal{F}_t)_{t\in[0,T]}$ for which we have $\mathbb{E}\sup_{t\le T}|X_t|^2 < \infty$. We set*

$$\|X\|_2 := [\mathbb{E} \sup_{t\le T} |X_t|^2]^{1/2}.$$

Then, upon identifying indistinguishable processes, $\|\cdot\|_2$ defines a norm. Furthermore, equipped with $\|\cdot\|_2$, $\mathcal{M}^{2,c}$ is a Banach space.

Remark 3.18. Note that for a function $f : [a, b] \to \mathbb{R}$, the mapping $t \mapsto \sup_{t\in[a,b]} f(t)$ is convex, so Jensen's inequality implies that for a stochastic process X as in the lemma we have

$$\sup_{t\leq T}\Big(\mathbb{E}|X_t|^2\Big) \leq \mathbb{E}\Big(\sup_{t\leq T}|X_t|^2\Big) < \infty.$$

Hence any process in $\mathcal{M}^{2,c}$ is also square-integrable over $[0, T]$ with respect to P. On the other hand, if X_t is a martingale, and $\sup_{t\leq T}\Big(\mathbb{E}|X_t|^2\Big) < \infty$, then by Lemma 2.15 we have that $|X_t|^2$ is a submartingale, so then Lemma 2.21 (Doob's L^p inequality) for $p = 2$ implies, that

$$\mathbb{E}\sup_{t\leq T}|X_t|^2 \leq 4\sup_{t\leq T}\mathbb{E}|X_t|^2 < \infty.$$

Hence when X_t is a martingale we could alternatively consider the equivalent norm defined by the mapping $X_t \mapsto \Big(\sup_{t\leq T}\mathbb{E}|X_t|^2\Big)^{\frac{1}{2}}$.

Proof of Lemma 3.17. Note that the fact that $\mathcal{M}^{2,c}$ is indeed a vector space follows from the linearity of the conditional expectation on one hand, and from the simple fact from real analysis that measurable functions form a vector space. Next, in order to verify that $\|\cdot\|_2$ indeed defines a norm it suffices to note that for a fixed ω, the map $X_t \mapsto \sup_t|X_t|$ defined on the space of real-valued continuous functions is a norm, and likewise for a fixed $t \in \mathbb{R}_+$, the mapping $\sup_t X_t \mapsto (E|\sup_t X_t|^2)^{\frac{1}{2}}$ is the L^2-norm on the space of functions $f : \mathbb{R} \to \mathbb{R}$. Thus for $X_t, Y_t \in \mathcal{M}^{2,c}$, we verify the triangle inequality by repeatedly using the fact that the aforementioned two maps are norms on their own right:

$$\begin{aligned}
\sup_{t\leq T}|X_t + Y_t| &\leq \sup_{t\leq T}|X_t| + \sup_{t\leq T}|Y_t| \Rightarrow \\
(\mathbb{E}(\sup_{t\leq T}|X_t + Y_t|)^2)^{\frac{1}{2}} &\leq (\mathbb{E}(\sup_{t\leq T}|X_t| + \sup_{t\leq T}|Y_t|)^2)^{\frac{1}{2}} \\
&\leq (\mathbb{E}(\sup_{t\leq T}|X_t|^2)^{\frac{1}{2}} + (\mathbb{E}(\sup_{t\leq T}|Y_t|^2)^{\frac{1}{2}}.
\end{aligned}$$

The fact that $\|\cdot\|_2$ is absolutely homogeneous is obvious. As alluded in the statement of the lemma, we can turn $\|\cdot\|_2$ into a norm by identifying indistinguishable processes, so that if $\|X\|_2 = 0$, then $X = 0$ holds as well, and as a result $\|\cdot\|_2$ indeed is a norm.
To prove that the space $\mathcal{M}^{2,c}$ is complete, consider a fixed $t \in [0, T]$, and let $\{X_t^{(n)}\}_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathcal{M}^{2,c}$. Note that by Theorem 2.21, namely Doob's L^p Inequal-

ity with $p = 2$ this sequence is convergent in the $L^2(P)$-norm:

$$\mathbb{E} \sup_{t \leq T} |X_t^{(n)} - X_t^{(m)}|^2 \leq 4 \sup_{t \leq T} \mathbb{E} |X_t^{(n)} - X_t^{(m)}|^2 \to 0, \text{ as } n, m \to \infty, \tag{3.19}$$

since the space $L^2(P)$ is complete. Hence there exists a limit for the sequence $\{X_t^{(n)}\}_{n \in \mathbb{N}}$ with respect to the $\| \cdot \|_2$-norm, denoted by M_t. To complete the proof, we must verify that M_t is a continuous martingale. Since $\{X_t^{(n)}\}_{n \in \mathbb{N}}$ converges to M_t in the $\| \cdot \|_2$-norm, which is equivalent to saying that

$$\mathbb{E} \sup_{t \leq T} |X_t^{(n)} - M_t|^2 \to 0, \text{ as } n \to \infty,$$

we are guaranteed a subsequence $\{n_k\}_{k=1}^{\infty}$ such that

$$\lim_{k \to \infty} \sup_{t \leq T} |X_t^{(n_k)} - M_t|^2 \to 0, \text{ almost surely on } \Omega.$$

This implies that

$$\lim_{k \to \infty} \sup_{t \leq T} |X_t^{(n_k)} - M_t| = 0, \text{ almost surely on } \Omega.$$

Note that above we have established that $\{X_t^{(n_k)}\}_{k=1}^{\infty}$ converges to M_t uniformly almost surely as $k \to \infty$. Since each $X_t^{(n_k)}$ is a continuous process, then up to indistinguishability, as a uniform limit of continuous functions, the function $t \mapsto M_t$ is continuous.
Finally, to show that M_t is a martingale, we verify that Definition 2.14 is satisfied. Note that we obtained M_t as a uniform limit of martingales $\{X_t^{(n_k)}\}$, except in set of P-measure zero. Measurability is preserved under limits, and as the filtration $(\mathcal{F}_t)_{t \in [0,T]}$ is complete, thus containing every P-null set by definition, we have that the limit process M_t is adapted.
Next, let $s \leq t$. We want to show that $M_s = \mathbb{E}_{\mathcal{F}_s} M_t$, or equivalently, that for any $A \in \mathcal{F}_s$, we have

$$\int_A M_s dP = \int_A M_t dP.$$

Note that by the earlier remark on equivalence of the norms on $\mathcal{M}^{2,c}$, we have that $X_t^{(n)} \to M_t$ in $L^2(P)$ as $n \to \infty$. Hence we can use the Dominated convergence theorem (which also guarantees that M_t is integrable since $L^2(\Omega) \subset L^1(\Omega)$ by the Cauchy-Schwartz inequality), in addition to the fact that each $X_t^{(n)}$ is a martingale to compute:

$$\int_A M_s dP = \lim_{n \to \infty} \int_A X_s^{(n)} dP = \lim_{n \to \infty} \int_A X_t^{(n)} dP = \int_A M_t dP,$$

which shows that M_t is a martingale, and the proof is finished. □

Remark 3.20. In particular Lemma 3.17 is well-suited for generalization of the Itô integral with respect to other martingales than just Brownian motion. Although our filtration $(\mathcal{F}_t)_{t\in[0,T]}$ has been chosen with the Brownian motion in mind, still nothing on the proof hinges upon the properties of the stochastic process W_t itself.

Building upon this remark, we now put Lemma 3.17 into good use, along with the properties of the Itô integral as constructed in Section 3.1. Let $\{X_t^{(n)}\}_{n\in\mathbb{N}}$ be a Cauchy sequence of stochastic processes as in (3.14), converging in $L^2(dt\otimes dP)$. For any $Y\in L_0$, the Itô integral $I_t(Y)$ is a continuous, square-integrable martingale by Lemma 3.7. This enables us to use Doob's L^p Inequality 2.21 with $p=2$ and then the Itô isometry to show that:

$$
\begin{aligned}
\mathbb{E}\sup_{t\le T}|I_t(X^{(n)})-I_t(X^{(m)})|^2 &= \mathbb{E}\sup_{t\le T}|I_t(X^{(n)}-X^{(m)})|^2\\
&\le 4\sup_{t\le T}\mathbb{E}|I_t(X^{(n)}-X^{(m)})|^2\\
&= 4\sup_{t\le T}\int_0^t\mathbb{E}|X_s^{(n)}-X_s^{(m)}|^2ds\to 0,\ \text{as } n\to\infty,
\end{aligned}
\tag{3.21}
$$

since $L^2(dt\otimes dP)$ is complete. The above implies that $\{I_t(X^{(n)})\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{M}^{2,c}$, which in turn by Lemma 3.17 is complete, so there exists a square-integrable martingale, denoted by Z_t, for which it holds that $\sup_{t\le T}|Z_t-I_t(X^{(n)})|\to 0$, in $L^2(P)$ as $n\to\infty$. Also, note that from inequality (3.21) it follows easily that the convergence on the left-hand side is not dependent on the chosen representative of $\{X_t^{(n)}\}_{n\in\mathbb{N}}$. Hence we have shown that the following definition for Z_t is reasonable:

Definition 3.22. (Itô integral). Suppose $X\in L^2_{ad}$, and let $\{X^{(n)}\}_{n=1}^\infty\subset L_0$ for which

$$\mathbb{E}\int_0^T|X_t^{(n)}-X_t|^2dt\to 0,\quad \text{as } n\to\infty.$$

Then we define the Itô integral of X with respect to the Brownian motion W to be the limit

$$I_t(X):=\lim_{n\to\infty}I_t(X^{(n)}),\quad \text{in } L^2(P).$$

For simplicity and clarity, in the sequel we will reserve the notation $I_t(X)$ to be as in Definition 3.22, unless otherwise explicitly specified.

Example 3.23. As an example of a rather simple Itô integral we compute directly from the definition that:

$$\int_a^b W_t^2dW_t=\frac{1}{3}(W_b^3-W_a^3)-\int_a^b W_tdt. \tag{3.24}$$

As an aside, it should be noted that only in rare cases can integrals of this type evaluated explicitly, and typically as an application of the Itô formula introduced at the end of this section. Note also that from the point of view of the fundamental theorem of (deterministic) calculus, the third term on the right-hand side of the above equation seems extraneous. This 'correction term' too is explained by the aforementioned Itô formula.

Now in order to compute the integral, we must first show that the assumptions of Definition 3.22 are met. That is, we must find a sequence $\{W_t^{(n)}\}_{n\in\mathbb{N}} \subset L_0$ such that $|W_t^{(n)} - W_t| \to 0$, as $n \to \infty$ in $L^2(dt \otimes dP)$. To this end, for a mesh $\Delta_n = \{t_0, t_1, ..., t_{n-1}, t_n\}$ of $[a, b]$ we define

$$W^{(n)}(t,\omega) = \mathbb{1}_{|W|\leq n} W(t_{i-1},\omega), \quad \text{for } t \in [t_{i-1}, t_i].$$

Clearly $W_t^{(n)}$ is an adapted sequence. Now by (iv) of Definition 2.22 almost all sample paths of $W_t(\omega)$ are continuous, so $W^{(n)}(t,\omega) \to W(t,\omega)$, as $n \to \infty$, almost surely on Ω. Furthermore,

$$\left|W^{(n)}(t,\omega)\right| \leq \sup_{t\in[a,b]} |W(t,\omega)|.$$

By the arcsine law of the Brownian motion, due to Paul Lévy in [22], the supremum process of the Brownian motion follows the arcsine distribution, and thus $\sup_{t\in[a,b]} W_t(\omega)$ has a finite second moment. Hence by the dominated convergence theorem

$$\mathbb{E}|W_t^{(n)}(\omega) - W_t(\omega)|^2 \to 0, \quad \text{as } n \to \infty.$$

Now the basic inequality $(a+b)^2 \leq 2a^2 + 2b^2$, and also (ii) in Definition 2.22 yield

$$\mathbb{E}|W_t^{(n)} - W_t|^2 \leq 2\mathbb{E}|W_t^{(n)}|^2 + 2\mathbb{E}|W_t|^2 \leq 4 \sup_{t\in[a,b]} \mathbb{E}|W_t|^2 \leq 4b < \infty.$$

Thus we can apply the Lebesgue dominated convergence theorem for $L^1(dt)$, and conclude that

$$\lim_{n\to\infty} \int_a^b \mathbb{E}|W_t^{(n)} - W_t|^2 dt = 0,$$

which by the Fubini theorem implies that also

$$\lim_{n\to\infty} \mathbb{E} \int_a^b |W_t^{(n)} - W_t|^2 dt = 0,$$

so

$$\int_a^b W_t^2 dW_t = \lim_{n\to\infty} \sum_{i=1}^n W_{t_{i-1}}^2 (W_{t_i} - W_{t_{i-1}}),$$

indeed exists as a limit in $L^2(P)$.
Next, it is a simple computation to verify that

$$
\begin{aligned}
3\sum_{i=1}^{n} W_{t_i}^2(W_{t_i}-W_{t_{i-1}}) = W_b^3 - W_a^3 - \sum_i (W_{t_i}-W_{t_{i-1}})^3 \\
-3\sum_i W_{t_{i-1}}(W_{t_i}-W_{t_{i-1}})^2.
\end{aligned}
\tag{3.25}
$$

Note that for a Brownian motion the increment $W_t - W_s \sim N(0, t-s)$ by definition, so therefore by the properties of the normal distribution we have $\mathbb{E}|W_t - W_s|^6 = 15|t-s|^3$. Hence it follows that:

$$
\mathbb{E}\Big|\sum_i (W_{t_i}-W_{t_{i-1}})^3\Big|^2 = 15\sum_i (t_i - t_{i-1})^3 \le 15 \max_{1\le i\le n}(t_i - t_{i-1})^2(b-a) \to 0,
\tag{3.26}
$$

as $n \to \infty$. For the second summation in (3.25) we compute the expectation by conditioning. Indeed, first notice that simply writing out the squared sum we have

$$
\begin{aligned}
&\mathbb{E}\Big|\sum_i W_{t_{i-1}}(W_{t_i}-W_{t_{i-1}})^2 - W_{t_{i-1}}(t_i - t_{i-1})\Big|^2 \\
&= \mathbb{E}\sum_{i,j}\Big(W_{t_{i-1}}(W_{t_i}-W_{t_{i-1}})^2 - W_{t_{i-1}}(t_i - t_{i-1})\Big)\Big(W_{t_{j-1}}(W_{t_j}-W_{t_{j-1}})^2 - W_{t_{j-1}}(t_j - t_{j-1})\Big),
\end{aligned}
$$

but since W_t has independent increments by definition, so only those terms for which $i = j$ remain in the above double sum. Hence we have

$$
\begin{aligned}
&\mathbb{E}\Big|\sum_i W_{t_{i-1}}(W_{t_i}-W_{t_{i-1}})^2 - W_{t_{i-1}}(t_i - t_{i-1})\Big|^2 = \\
&\mathbb{E}\Big[\sum_i W_{t_{i-1}}^2(W_{t_i}-W_{t_{i-1}})^4 + W_{t_{i-1}}^2(t_i - t_{i-1})^2 - 2W_{t_{i-1}}^2(W_{t_i}-W_{t_{i-1}})^2(t_i - t_{i-1})\Big].
\end{aligned}
$$

Next, conditioning with respect to the filtration $\mathcal{F}_{t_{i-1}}$, respect to which $W_{t_{i-1}}$ is adapted, we may use property (v) in Proposition 2.13, and also $\mathbb{E}|W_t - W_s|^4 = 3|t-s|$, again by the properties of the normal distribution, to continue the above chain of equalities:

$$\begin{aligned}
&\mathbb{E}\Big[\sum_i W_{t_{i-1}}^2 (W_{t_i} - W_{t_{i-1}})^4 + W_{t_{i-1}}^2 (t_i - t_{i-1})^2 - 2W_{t_{i-1}}^2 (W_{t_i} - W_{t_{i-1}})^2 (t_i - t_{i-1})\Big]\\
=&\mathbb{E}\Big[\mathbb{E}_{\mathcal{F}_{t_{i-1}}} \sum_i W_{t_{i-1}}^2 (W_{t_i} - W_{t_{i-1}})^4 + W_{t_{i-1}}^2 (t_i - t_{i-1})^2 - 2W_{t_{i-1}}^2 (W_{t_i} - W_{t_{i-1}})^2 (t_i - t_{i-1})\Big]\\
=&\mathbb{E}\Big[\sum_i W_{t_{i-1}}^2 \cdot 3(t_i - t_{i-1})^2 + W_{t_{i-1}}^2 (t_i - t_{i-1})^2 - 2W_{t_{i-1}}^2 (t_i - t_{i-1})(t_i - t_i)\Big]\\
=&\sum_i 3t_{i-1}(t_i - t_{i-1})^2 + t_{i-1}(t_i - t_{i-1})^2 - 2(t_{i-1})(t_i - t_{i-1})^2\\
=&2\sum_i t_{i-1}(t_i - t_{i-1})^2 \leq 2b \max_{1\leq i\leq n} (t_i - t_{i-1})(b-a).
\end{aligned}$$

Hence it follows that

$$\begin{aligned}
&\mathbb{E}\Big[\sum_i W_{t_{i-1}}^2 (W_{t_i} - W_{t_{i-1}})^4 + W_{t_{i-1}}^2 (t_i - t_{i-1})^2 - 2W_{t_{i-1}}^2 (W_{t_i} - W_{t_{i-1}})^2 (t_i - t_{i-1})\Big]\\
&\qquad\leq 2b \max_{1\leq i\leq n} (t_i - t_{i-1})(b-a) \to 0,
\end{aligned} \tag{3.27}$$

as $n \to \infty$. This shows that

$$\lim_{n\to\infty} \sum_i W_{t_{i-1}} (W_{t_i} - W_{t_{i-1}})^2 = \int_a^b W_t dt, \quad \text{in } L^2(P).$$

Note also that (3.26) shows that

$$\lim_{n\to\infty} \sum_i (W_{t_i} - W_{t_{i-1}})^3 = 0,$$

in $L^2(P)$. Therefore combining equations (3.26) and (3.27) with equation (3.25), we have shown that equation (3.24) holds.

Lemma 3.28. *(Itô Isometry). Let W be a Brownian motion, and $X \in L^2_{ad}$. Then we have the Itô isometry*

$$\mathbb{E}(|(I_t(X)|^2) = \mathbb{E}\Big(\int_0^t X_s^2 ds\Big).$$

Proof. The claim is essentially that of Lemma 3.7, the Itô isometry for simple processes, adjusted for the proper setting of this section. The proof of the isometry follows directly from Lemma 3.7, by use of Lemma 3.17. Indeed, let $X \in L^2_{ad}$, and let $\{X^{(n)}\}_{n\in\mathbb{N}} \subset L_0$

be such that $X^{(n)} \to X$ in $L^2(dt \otimes dP)$ as $n \to \infty$. Now $I_t(X)$ is by definition the $L^2(P)$-limit of $\{I_t(X^{(n)})\}_{n\in\mathbb{N}}$, so the two aforementioned lemmas, Lemma 3.7, and Lemma 3.17, along with a repeated use of Lebesgue dominated convergence theorem yield:

$$\begin{aligned}\mathbb{E}\big(|I_t(X)|^2\big) &= \mathbb{E}\big(\lim_{n\to\infty}|I_t(X^{(n)})|^2\big) = \lim_{n\to\infty}\mathbb{E}\big(|I_t(X^{(n)}|^2\big)\\ &= \lim_{n\to\infty}\mathbb{E}\int_0^t |X_s^{(n)}|^2 ds\\ &= \mathbb{E}\int_0^t |X_s|^2 ds,\end{aligned}$$

which is precisely the claimed Itô isometry for $X \in L^2_{ad}$. □

Example 3.29. We continue with the Itô integral given in Example 3.23, and define a stochastic process on $[0,T]\times\Omega$ by setting

$$X_t := \int_0^t W_s^2 dW_s = \frac{1}{3}W_t^3 - \int_0^t W_s ds.$$

Now clearly X_t is $\mathcal{F}_t$-adapted. In addition, we claim that X_t is a martingale. Note that as $X_t \in L^2_{ad}$, then in particular X_t is P-integrable. Now, let $s \le t$, and we want to show that

$$\mathbb{E}_{\mathcal{F}_s} X_t = X_s,$$

and by equation (3.24) it is enough to verify that

$$\mathbb{E}_{\mathcal{F}_s}\Big[\frac{1}{3}W_t^3 - \int_0^t W_s ds\Big] = \frac{1}{3}W_s^3 - \int_0^s W_u du. \tag{3.30}$$

It is a simple computation to check the identity

$$W_t^3 = (W_t - W_s)^3 + 3W_s(W_t - W_s)^2 + 3W_s^2(W_t - W_s) + W_s^3,$$

and hence as a result, taking the conditional expectation of the above equation yields:

$$\begin{aligned}\mathbb{E}_{\mathcal{F}_s} W_t^3 &= \mathbb{E}_{\mathcal{F}_s}(W_t - W_s)^3 + 3\mathbb{E}_{\mathcal{F}_s} W_s(W_t - W_s)^2 + 3\mathbb{E}_{\mathcal{F}_s} W_s^2(W_t - W_s)\\ &\quad + \mathbb{E}_{\mathcal{F}_s} W_s^3\\ &= 0 + W_s\mathbb{E}_{\mathcal{F}_s}(W_s - W_s)^2 + 3W_s^2\mathbb{E}_{\mathcal{F}_s}(W_t - W_s) + \mathbb{E}_{\mathcal{F}_s} W_s^3\\ &= W_s(t-s) + W_s^3, \qquad (3.31)\end{aligned}$$

where we have used the following facts: since W_s is $\mathcal{F}_s$-adapted, $W_t - W_s \sim N(0, t-s)$, and furthermore has independent increments, and thus properties (vi), (vii) of Proposition 2.13 are applicable. Also, we can write

$$\mathbb{E}_{\mathcal{F}_s}\Big(\int_0^t W_u du\Big) = \mathbb{E}_{\mathcal{F}_s}\Big(\int_0^s W_u du + \int_s^t W_u du\Big),$$

where clearly $\int_0^s W_u du$ is $\mathcal{F}_s$-adapted. Hence by the linearity of the conditional expectation and (vi) in Proposition 2.13, with the use of the Fubini theorem, we have:

$$\begin{aligned}\mathbb{E}_{\mathcal{F}_s}\Big(\int_0^t W_u du\Big) &= \int_0^s W_u du + \int_s^t (\mathbb{E}_{\mathcal{F}_s} W_u) du \\ &= \int_0^s W_u du + \int_s^t W_s du \\ &= \int_0^s W_u du + W_s(t-s).\end{aligned} \tag{3.32}$$

Where the second equality follows from the fact that W_t is a martingale, namely Proposition 2.24. Now equations (3.31) and (3.32) together imply that (3.30) holds, and thus X_t is indeed a martingale.

Note that since all sample paths of W_t are almost surely continuous, it follows that the stochastic process X_t defined in the previous example belongs to the space L^2_{ad}. The next proposition shows that the martingale property holds true for any Itô integral for $X \in L^2_{ad}$:

Proposition 3.33. *Let W be a Brownian motion, and $X \in L^2_{ad}$. Then the Itô integral of X is a square-integrable martingale with respect to the filtration $(\mathcal{F}_t)_{t\in[0,T]}$.*

Proof. Note that the square-integrability of $I_t(X)$ follows directly from Lemma 3.28. In order to finish the proof, it is enough to note that proof of the fact that $I_t(X)$ is a martingale follows from the end of Lemma 3.17. □

Remark 3.34. One way to work around problems with martingales is to consider the property only locally. We define a stochastic process X_t to be a *local martingale* with respect to a filtration $(\mathcal{F}_t)_{t\in[0,T]}$ if there exists a sequence of stopping times $\{\tau_n\}_{n\in\mathbb{N}}$ such that $\tau_n \to \infty$ almost surely as $n \to \infty$, and for each $n \in \mathbb{N}$ the stopped process $X_{t\wedge\tau_n}$ is a martingale. Note that in contrast to a martingale, a priori a local martingale X_t may not be P-integrable. Naturally each martingale is a local martingale. The concept of a local martingale was first considered by Kiyosi Itô and Shinzo Watanabe in [16]. Localization techniques are widely used in stochastic analysis, see for instance [7], [35] or [34].

Proposition 3.35. *Let W be a Brownian motion, and $X \in L^2_{ad}$. Then for any $t \in [0,T]$, the Itô integral is a random variable with zero mathematical expectation, that is, $\mathbb{E}(I_t(X)) = 0$.*

Proof. By the previous proposition, $I_t(X)$ is a martingale, and we already noted in (ii) of Proposition 3.5 that $I_0(X) = 0$. Hence by using property (v) in Proposition 2.13 and conditioning with respect to the sigma-algebra $\mathcal{F}_0$, we have:

$$\mathbb{E} I_t(X) = \mathbb{E}\Big[\mathbb{E}_{\mathcal{F}_0} I_t(X)\Big] = \mathbb{E}0 = 0.$$

The proof is finished. □

Theorem 3.36. *(Continuity property of the Itô integral). Suppose $X \in L^2_{ad}$, and let $t \in [0,T]$. Then the stochastic process given by*

$$Y_t = \int_0^t X_s dW_s,$$

is continuous, that is, almost all of its sample paths are continuous functions on $[0,T]$.

Proof. For clarity, we will rephrase the relevant part proof of Lemma 3.17: By the definition of the Itô integral, we have that $I_t(X^{(n)}) \to I_t(X)$ in $L^2(P)$, and hence there exists a subsequence $\{n_k\}$ such that the sequence

$$\sup_{t \leq T} |I_t(X^{(n_k)}) - I_t(X)| \to 0, \text{ almost surely on } \Omega,$$

as $k \to \infty$. As each $I_t(X^{(n)})$ is almost surely continuous by Lemma 3.7, the uniform limit of this sequence except in a set of P-measure zero, $I_t(X)$, is almost surely a continuous function as well, and the claim is proved. □

In this section we have defined the Itô integral for a suitable class of processes, and established some of its key properties. As Theorem 3.36 shows, many of these have their counterpart in Lebesgue integration theory, and in light of Lemma 3.28, this is not tremendously surprising. There are notable differences however. For instance the Itô integral is not monotone in the sense that if $X_t \leq Y_t$ almost surely, we have $I_t(X) \leq I_t(Y)$ as well. This is shown by the following example:

Example 3.37. Let $s \in [0,T]$, and let $X_t \equiv 0$, and let $Y_t := \mathbb{1}_{[0,s]}(t)$, for all $t \in [0,T]$. Now clearly $X_t \leq Y_t$, and it is also clear that as constants $X_t, Y_t \in L^2_{ad}$. From the definition of the Itô integral we have immediately that $I_t(X) = 0$, but

$$I_t(Y) = \int_0^t \mathbb{1}_{[0,s]}(u) dW_u = \int_0^{s \wedge t} 1 \cdot dW_u = W_{s \wedge t}.$$

It is well-known (see, for instance [10]) that for a Brownian motion W we have almost surely for every $\epsilon > 0$, that $\inf_{0 \leq s \leq \epsilon} W_s < 0$, thus contradicting $I_t(X) \leq I_t(Y)$.

To conclude, we briefly revisit the point hinted at in Remark 3.4, namely that of the other possible choices of integrators beside the Brownian motion W. A natural extension in view of Theorem 3.17 is to consider square-integrable martingales M in general, and extend the martingale transform $(X \cdot M)_t$. To proceed, we define the quadratic variation of a continuous square-integrable martingale M_t to be the continuous, adapted increasing process denoted by $\langle M \rangle_t$ for which $M_t^2 - \langle M \rangle_t$ is a martingale, and $\langle M \rangle_0 = 0$. It is a fact that for every continuous square-integrable martingale M_t such $\langle M \rangle_t$ indeed exists, see for instance [3]. For example, on the bounded interval $[0, T]$ we have for the Brownian motion that $\langle W \rangle_t = t$.
Hence it seems reasonable to require in Definition 3.22 that $\mathbb{E} \int_0^T X_s^2 d\langle M \rangle_s < \infty$, and then it can be shown that the cornerstone of the resulting integration theory, the Itô isometry in Lemma 3.28 becomes

$$\mathbb{E}|I_t(X)|^2 = \mathbb{E} \int_0^T X_s^2 d\langle M \rangle_s.$$

Yet another popular alternative is to consider integration with respect to semimartingales, as done in [32] or [7], for instance. A starting point is to define that an adapted process A_t is a finite variation process if all its sample paths are finite variation functions on $[0, \infty)$. Then we define a stochastic process Y_t to be a continuous semimartingale if it can be written as a sum

$$Y_t = M_t + A_t,$$

where M_t is a continuous local martingale, and A_t is a finite variation process. Now

$$\int_0^t X_s dY_s$$

can be defined via the decomposition of Y as follows. First, we define the general Itô integral $\int_0^{t \wedge \tau_n} X_s dM_s$ by the construction of the previous paragraph, with $\{\tau_n\}_{n \in \mathbb{N}}$ being the sequence of stopping times from the definition of the local martingale M_t. Second, since A_t is of finite variation, $\int_0^t X_s dA_s$ can be defined as a Lebesgue-Stieltjes integral. Hence we define $\int_0^t X_s dY_s$ to be the sum of these two, and it can be shown that this intuitive definition is reasonable as well.
We now expand on Example 3.23, where we computed that

$$\int_a^b W_t^2 dW_t = \frac{1}{3}\left(W_b^3 - W_a^3\right) - \int_a^b W_t dt.$$

A reasonably easy way to see how the stochastic integration theory developed above differs from the classical Leibniz-Newton calculus is to consider the chain rule. In the latter, for two differentiable functions f and g we have $\frac{d}{dt}f(g(t)) = f'(g(t))g'(t)$ which can be equivalently written as $f(g(t)) - f(g(a)) = \int_a^t f'(g(s))g'(s)ds$ via the fundamental theorem of calculus. However, in the case of the Itô integral, we have, for a twice continuously differentiable function f the famous Itô formula, given by the next theorem: .

Theorem 3.38. *(Itô formula, first version). Let W be a Brownian motion, and let $f : \mathbb{R} \to \mathbb{R}$ be of class C^2. Then the following Itô formula holds:*

$$f(W_t) = f(W_a) + \int_a^t f'(W_s)dW_s + \frac{1}{2}\int_a^t f''(W_s)ds. \tag{3.39}$$

Proof. Omitted. See, for instance [21]. □

Notice that in equation (3.39) the second term on the right is an Itô integral as given in Definition 3.22, and the third is an ordinary Riemann integral, heuristically corresponding to the nonzero quadratic variation of the Brownian motion. Equation 3.39 is often written in differential notation as

$$df(W_t) = f'(W_t)dW_t + \frac{1}{2}f''(W_t)dt,$$

but this differential form on its own implies nothing more than the corresponding integral equation, (3.39). To briefly expand on this matter, note that since for all $t \geq 0$ the Brownian motion process W_t is non-differentiable [24], we can instead define the derivative of W_t, called the white noise, by the relation

$$\int_0^\infty g(t)W_t'dt = -\int_0^\infty g'(t)W_t dt, \text{ for all } g \in C_0^\infty.$$

In other words W' is defind as the weak (distributional) derivative of W_t. For details of this distributional approach, see for instance [28].
It can be shown that a slightly more general version of (3.39) holds:

Theorem 3.40. *(Itô formula, second version). Let W be a Brownian motion, and let X be a stochastic process of the form*

$$dX_t = f(t)dW_t + g(t)dt,$$

where both f and g are adapted to the filtration generated by W and $f \in L^2([a,b] \times \Omega)$ and $g \in L^1(a,b)$. Let $\theta(t,x)$ be a continuous function, with $\frac{\partial\theta}{\partial t}$, $\frac{\partial\theta}{\partial x}$ and $\frac{\partial^2\theta}{\partial x^2}$ also continuous. Then for the process $\theta(t, X_t)$ the generalized Itô formula holds:

$$d\theta(t,X_t) = \frac{\partial\theta}{\partial t}(t,X_t)dt + \frac{\partial\theta}{\partial x}(t,X_t)dX_t + \frac{1}{2}\frac{\partial^2\theta}{\partial x^2}(t,X_t)(dX_t)^2, \tag{3.41}$$

where $(dX_t)^2 = dX_t \cdot dX_t$ *is computed according to the rule* $dt \cdot dt = dt \cdot dW_t = dW_t \cdot dt = 0$*, and* $dW_t \cdot dW_t = dt$*.*

Proof. Omitted. See, for instance [17], [30] or [34]. □

A vital property of the Itô integral as developed in this chapter is given by Lemma 3.36, namely that the Itô integral is a martingale. This allows us to use the well-developed theoretical toolbox of martingale theory, and indeed much of the proof of the main theorem in the next chapter, Theorem 4.13, exploits this connection extensively. In fact, two separate observations can be made of this matter.
First, as alluded earlier in Remark 3.3, the point where the integrand ξ is evaluated at is chosen specifically to guarantee that the resulting $L^2(P)$-limit is a martingale. In fact, instead of choosing the left endpoint of the interval (t_{i-1}, t_i), we could use $t_i^* = \frac{t_{i-1}+t_i}{2}$, and the limit of the resulting sum, called the *Stratonovich integral*, would still converge in $L^2(P)$, however the martingale property would be lost. This type of integral was first introduced by Ruslan Stratonovich in [36], and is often used in applications such as signal processing. [7] For an overview of its main properties and contrasting features with the Itô integral, the interested reader may consult [21], [7] or [43].
Secondly, in order to guarantee the martingale property of the Itô integral $I_t(X)$, the class of integrands $X \in L^2_{ad}$ was chosen to be suitably small. It is indeed possible to define the Itô integral for a larger class of integrands by relaxing the integrability requirement: instead of requiring that $\mathbb{E} \int_0^T |X_t(\omega)|^2 dt < \infty$, we can insist that we only have $\int_0^T |X_t(\omega)|^2 dt < \infty$ almost surely. Then the resulting Itô integral is no longer necessary P-integrable, and therefore not a martingale (but a local martingale) and also the Itô isometry is not preserved. For details of such an approach, we refer to [21] or [35].

Chapter 4

Stochastic Differential Equations

4.1 Definitions and some examples

Definition 4.1. Let W be a Brownian motion and ξ be a random variable independent of W on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,\infty)}, P)$, a complete filtered probability space equipped with a filtration generated by ξ and the natural filtration of W, satisfying the usual conditions. An adapted stochastic process X is a solution of a stochastic differential equation (SDE) if we have

$$X_t = \xi + \int_0^t \sigma(s, X_s) dW_s + \int_0^t f(s, X_s) ds \tag{4.2}$$

almost surely for all $t \in \mathbb{R}_+$. The functions σ and f are called the coefficients of the equation. Note that if $\sigma \equiv 0$, then (4.2) reduces to an ordinary differential equation.

Remark 4.3. Note that in the above definition we a priori require that the random variables ξ and W are independent. Via a monotone class argument, this guarantees that W remains a Brownian motion with respect to the filtration generated by ξ and the natural augmented filtration of W.

In view of the discussion in the end of the Chapter 3, we may write equation (4.2) in the form of a stochastic initial-value problem

$$dX_t = \sigma(t, X_t)dW_t + f(t, X_t)dt, \quad X_0 = \xi.$$

Although illustrative of the term stochastic *differential* equation, it is worth bearing in mind that this initial-value problem form can be deceptive. Partly this is due to the reasons discussed earlier, but also note that in general ξ is considered to be a random variable on Ω, not simply a fixed real number. Hence the *stochastic* in the stochastic

differential equation (4.2) is not solely limited to the Itô integral term, although admittedly it is the major theoretical stumbling block.
Before proceeding with the main theorem of this chapter, we give two examples from [21] in order to demonstrate the challenges concerning SDE existence and uniqueness. Note that as the author in [21] points out, these are very similar in structure to the standard examples in ordinary differential equations literature.

Example 4.4. (Example of no global solution). Consider the stochastic differential equation
$$dX_t = X_t^2 dW_t + X_t^3 dt, \quad X_0 = 1.$$
We solve this using the generalized Itô formula (3.41). Let $\theta(t, X_t) = \frac{1}{X_t}$. Note that Itô formula in equation (3.41) yields:
$$d\Big(\frac{1}{X_t}\Big) = -\frac{1}{X_t^2} dX_t + \frac{1}{2}\frac{2}{X_t^3}\Big(dX_t\Big)^2,$$
and inserting the given expression for dX_t, we compute (recalling the arithmetic $dt \cdot dt = dt \cdot dW_t = dW_t \cdot dt = 0$, $dW_t \cdot dW_t = dt$):
$$\begin{aligned} d\Big(\frac{1}{X_t}\Big) &= -\frac{1}{X_t^2}\Big(X_t^2 dW_t + X_t^3 dt\Big) + \frac{1}{X_t^3}\Big(X_t^2 dW_t + X_t^3 dt\Big)^2 \\ &= -dW_t - X_t dt + X_t dt \\ &= -dW_t, \end{aligned}$$
or equivalently $\frac{1}{X_t} = C - W_t$, with C a constant. The initial condition $X_0 = 1$ implies, since $W_0 = 0$ almost surely, that
$$X_t = \frac{1}{1 - W_t}.$$
Note that the solution is defined only for those $t \in \mathbb{R}_+$ for which the Brownian motion W remains in the interval $(-\infty, 1)$, and thus can (and in fact certainly will, see [2]) explode in finite time.

Example 4.5. (Example of infinitely many solutions). Consider the stochastic differential equation
$$dX_t = 3X_t^{2/3} dW_t + 3X_t^{1/3} dt, \quad X_0 = 0.$$
Define a function $\theta_a(x) = (x - a)^3 \mathbb{1}_{\{x \geq a\}}$ with constant $a > 0$ fixed. Now by Itô formula (3.39) we have
$$d\theta(W_t) = \theta'(W_t) dW_t + \frac{1}{2}\theta''(W_t) dt = 3(\theta_a(W_t))^{\frac{2}{3}} dW_t + 3\theta_a(W_t)^{\frac{1}{3}} dt,$$

since $\theta'(x) = 3\theta_a(x)^{\frac{2}{3}}$, and $\theta''(x) = 6\theta_a(x)^{\frac{1}{3}}$. Also we have $\theta_a(W_0)) = 0$. Thus for any $a > 0$, the given stochastic differential equation has a solution $\theta_a(W_t)$.

Motivated by these examples, we make the following definitions:

Definition 4.6. A measurable function $g(t,x)$ on $[a,b] \times \mathbb{R}$ satisfies the *Lipschitz condition in* x if there exists a constant $K > 0$ such that

$$|g(t,x) - g(t,y)| \leq K|x-y|,$$

for all $t \in [a,b], x, y \in \mathbb{R}$.

Definition 4.7. A measurable function $g(t,x)$ on $[a,b] \times \mathbb{R}$ satisfies the *linear growth condition in* x if there exists a constant $K > 0$ such that

$$|g(t,x)|^2 \leq K(1+x^2),$$

for all $t \in [a,b], x \in \mathbb{R}$.

Lemma 4.8. *(Bellman-Grönwall-inequality). Suppose that $g \in L^1(a,b)$ satisfies*

$$g(t) \leq f(t) + \beta \int_a^b g(s)ds, \tag{4.9}$$

for all $t \in [a,b]$, where $f \in L^1(a,b)$ and $\beta > 0$. Then

$$g(t) \leq f(t) + \beta \int_a^t f(s)e^{\beta(t-s)}ds.$$

In particular, when $f \equiv \alpha \in \mathbb{R}$, we have $g(t) \leq \alpha e^{\beta(t-a)}$, for all $t \in [a,b]$.

Proof. Define a new function $h(t)$ by

$$h(t) := \beta \int_a^t g(s)ds,$$

for $t \in [a,b]$. Now by the fundamental theorem of calculus and inequality (4.9) we have for almost every $t \in [a,b]$:

$$h'(t) = \beta g(t) \leq \beta f(t) + \beta h(t).$$

Note that this is the same as

$$\frac{d}{dt}(e^{-\beta t}h(t)) \leq \beta f(t)e^{-\beta t},$$

and integrating this equation from a to t yields

$$e^{-\beta t}h(t) \le \beta \int_a^t f(s)e^{-\beta s}ds \Leftrightarrow h(t) \le \int_a^t f(s)e^{\beta(t-s)}ds.$$

Thus by equation (4.9) we have

$$g(t) \le f(t) + h(t) \le f(t) + \beta \int_a^t f(s)e^{\beta(t-s)}ds,$$

as claimed. □

Lemma 4.10. *Let $\{h_n\}_{n=1}^{\infty} \subset L^1(a,b)$ satisfying*

$$h_{n+1}(t) \le f(t) + \beta \int_a^t h_n(s)ds, \tag{4.11}$$

for all $t \in [a,b]$, where $f \in L^1(a,b)$ and $\beta \in \mathbb{R}_+$. Then we have

$$h_{n+1}(t) \le f(t) + \beta \int_a^t f(u)e^{\beta(t-u)}du + \beta^n \int_a^t \frac{(t-u)^{n-1}}{(n-1)!}h_1(u)du,$$

for any $n \ge 1$. In particular, when $f \equiv \alpha$ and $h_1 \equiv c$ are constants, then the following inequality holds for any $n \ge 1$:

$$h_{n+1}(t) \le \alpha e^{b(t-a)} + c\frac{\beta^n(t-a)^n}{n!}.$$

Proof. To prove the claim, we will first show by induction that for any $n \ge 1$ we have

$$\begin{aligned} h_{n+1}(t) \le{} & f(t) + \beta \int_a^t f(s)ds + \beta^2 \int_a^t (t-u)f(u)du \\ & + \cdots + \beta^{n-1} \int_a^t \frac{(t-u)^{n-2}}{(n-2)!}f(u)du + \beta^n \int_a^t \frac{(t-u)^{n-1}}{(n-1)!}h_1(u)du. \end{aligned} \tag{4.12}$$

Note that the base case $n = 1$ follows immediately from the previous lemma since putting $n = 1$ in equation (4.11) we have

$$h_2(t) \le f(t) + \beta \int_a^t h_1(s)ds,$$

which is exactly the antecedent in the statement of Lemma 4.8. So $n = 1$ is clear. Then, assume that for some $n = k \in \mathbb{N}$ inequality (4.12) holds. We want to show it holds for $k+1$ as well. By equation (4.11) for $k+1$ we have the inequality

$$h_{k+2}(t) \le f(t) + \beta \int_a^t h_{k+1}(s)ds,$$

and using the inductive hypothesis for h_{k+1} in equation (4.12) we have

$$h_{k+2} \leq f(t) + \beta \int_a^t \Big[f(s) + \beta \int_a^s f(y)dy + \cdots + \beta^{k-1} \int_a^s \frac{(s-u)^{k-2}}{(k-2)!} f(u)du$$
$$+ \beta^k \int_a^s \frac{(s-u)^{k-1}}{(k-1)!} h_1(u)du \Big] ds.$$

Note that in the above for the general term we have, upon changing the order of integration:

$$\begin{aligned}
\beta \int_a^t \beta^k \int_a^s \frac{(s-u)^{k-1}}{(k-1)!} h_1(u)du\, ds &= \beta^{k+1} \int_a^t h_1(u) \int_u^t \frac{(s-u)^{k-1}}{(k-1)!} ds\, du \\
&= \beta^{k+1} \int_a^t h_1(u) \frac{(t-u)^k}{k(k-1)!} du \\
&= \beta^{k+1} \int_a^t \frac{(t-u)^k}{k!} h_1(u)du.
\end{aligned}$$

This is precisely the claimed general term for h_{k+2}. Thus by the principle of mathematical induction, formula (4.12) is valid for any $n \geq 1$. Next, note that inequality (4.12) can be simplified by the obvious estimate $\sum_{k=0}^{n-2} \beta^k \frac{(t-u)^k}{k!} \leq e^{\beta(t-u)}$. So we have

$$h_{n+1}(t) \leq f(t) + \beta \int_a^t f(u) e^{\beta(t-u)} du + \beta^n \int_a^t \frac{(t-u)^{n-1}}{(n-1)!} h_1(u)du,$$

and the proof is finished. □

4.2 Existence and uniqueness of solutions for equations with Lipschitz coefficients

Theorem 4.13. *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,\infty)}, P)$, be a complete filtered probability space as in Definition 4.1, and let W be a Brownian motion defined on this space. Suppose that ξ is a random variable on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,\infty)}, P)$, for which $\mathbb{E}\xi^2 < \infty$ and also that ξ is independent of the Brownian motion W. Suppose that the coefficients $\sigma(t,x), f(t,x)$ satisfy the global Lipschitz and linear growth conditions in x.*
Then there exists a continuous, $\mathcal{F}_t$-adapted process $(X_t)_{0\leq t<\infty}$ which is the unique solution of (4.2) *satisfying the initial condition ξ.*

The structure of the proof presented here is mainly adapted from [21], [30] and [31], a traditional approach first pioneered by Itô himself [15] in the 1940's. An alternative proof using the Banach Fixed Point Theorem can be found for instance in [32].
We first prove the existence and uniqueness result for $t \in [0, T]$ with $T > 0$ arbitrary, and then extend this result from compact intervals to the unbounded case. For notational simplicity we may assume that each of the global Lipschitz and linear growth condition constants are dominated by a single non-negative constant, denoted by C.

Proof of Theorem 4.13.

- The stochastic integral equation (4.2) has at most one continuous solution X_t on $[0, T]$.

  We argue by contradiction: assume that there exists two continuous solutions satisfying equation (4.2), X and Y, and let $Z_t := X_t - Y_t$. Then Z_t is clearly a continuous process, and by definition

$$Z_t = \int_0^t (\sigma(s, X_s) - \sigma(s, Y_s))dW_s + \int_0^t (f(s, X_s) - f(s, Y_s))ds.$$

  Now using the elementary inequality $(a + b)^2 \leq 2(a^2 + b^2)$ we have

$$Z_t^2 \leq 2\Big[\Big(\int_0^t (\sigma(s, X_s) - \sigma(s, Y_s))dW_s\Big)^2 + \Big(\int_0^t (f(s, X_s) - f(s, Y_s))ds\Big)^2\Big], \tag{4.14}$$

  and we can estimate the expectation of the first term on the right by

$$\begin{aligned} \mathbb{E}\Big(2\int_0^t (\sigma(s, X_s) - \sigma(s, Y_s))dW_s\Big)^2 &= 2\int_0^t \mathbb{E}\Big(\sigma(s, X_s) - \sigma(s, Y_s)\Big)^2 ds \\ &\leq 2\int_0^t \mathbb{E}\Big(C^2|X_s - Y_s|^2\Big)ds \\ &= 2C^2 \int_0^t \mathbb{E}|Z_s|^2 ds, \qquad (4.15) \end{aligned}$$

  where the first equality is an application of Theorem 3.28 (Itô isometry), and the inequality follows from the Lipschitz continuity of the coefficient $b(t, x)$. Now similarly for the second term on the right of equation (4.14) we have

$$\begin{aligned} 2\Big(\int_0^t (f(s, X_s) - f(s, Y_s))ds\Big)^2 &\leq 2\int_0^t (f(s, X_s) - f(s, Y_s))^2 ds \\ &\leq 2C^2 \int_0^t |Z_s|^2 ds, \end{aligned}$$

where we have first used the Jensen's inequality in the first inequality and then the Lipschitz continuity of $f(t,x)$ in the second. So the above implies

$$\mathbb{E}2\Big(\int_0^t (f(s,X_s) - f(s,Y_s))ds\Big)^2 \leq 2C^2 \int_0^t \mathbb{E}|Z_s|^2 ds, \tag{4.16}$$

where we have used Fubini's theorem to bring the expectation inside the integral. Note that this is justified due to Theorem 3.28.

Taking expectation on both sides of equation (4.14) and inserting the estimates in equations (4.15) and (4.16) we have:

$$\mathbb{E}(Z_t^2) \leq 4C^2 \int_0^t \mathbb{E}|Z_s|^2 ds,$$

and so by Lemma 4.8 (Bellman-Grönwall) we have $\mathbb{E}Z_t^2 \leq 0$ for all $t \in [0,T]$. But this is possible only when $\mathbb{E}Z_t^2 = 0$ for all $t \in [0,T]$. Now by basic properties of the integral this implies $Z_t = 0$ almost surely for all $t \in [0,T]$, that is, $X_t = Y_t$ almost surely for all $t \in [0,T]$. Since both X and Y were assumed to be continuous, in particular right-continuous, by Lemma 2.4 they are indistinguishable, and the uniqueness claim is proved.

- The stochastic differential equation (4.2) has a continuous solution X_t on $[0,T]$.

Define a sequence $\{X_t^{(n)}\}_{n\in\mathbb{N}}$ of stochastic processes inductively by setting $X_t^{(1)} = \xi$ and for $n \geq 1$:

$$X_t^{(n+1)} = \xi + \int_0^t \sigma(s, X_s^{(n)})dW_s + \int_0^t f(s, X_s^{(n)})ds. \tag{4.17}$$

First, we must show that $\{X_t^{(n)}\}_{n\in\mathbb{N}} \subset L^2_{ad}$ for the Itô integral to be well-defined. We proceed by induction.

We have that $X_t^{(1)} = \xi$ is clearly continuous in t. Also we check that $X_t^{(1)} \in L^2_{ad}$: indeed, it is immediate from the definition of a generated σ-algebra that for two collections $\mathcal{H},\mathcal{G}$ such that $\mathcal{H} \subset \mathcal{G}$, we have $\sigma(\mathcal{H}) \subset \sigma(\mathcal{G})$. As the filtration $(\mathcal{F}_t)_{t\in[0,T]}$ is generated by ξ and the natural filtration of W, by the previous remark $X_t^{(1)} = \xi$ is $\mathcal{F}_t$-adapted. Since $\mathbb{E}\xi^2 < \infty$ by assumption, we have that $X_t^{(1)} \in L^2_{ad}$.

Also for $n \geq 1$, each $X_t^{(n+1)}$ is continuous: the continuity of the Itô integral term follows from Theorem 3.36, and the Lebesgue integral term from basic real analysis.

Now assume that for some $n \geq 1$, $X_t^{(n)} \in L^2_{ad}$. Then, using the basic inequality $|a+b+c|^2 \leq 3(a^2+b^2+c^2)$ we have

$$|X_t^{(n+1)}|^2 \leq 3\Big[\xi^2 + (\int_0^b \sigma(s, X_s^{(n)})dW_s)^2 + (\int_a^b f(s, X_s^{(n)})ds)^2\Big]. \tag{4.18}$$

Then, using first the Itô isometry, available via the inductive hypothesis for $X_t^{(n)}$, and since $\sigma(t,x)$ satisfies the linear growth condition, we have:

$$\mathbb{E}(\int_0^b \sigma(s, X_s^{(n)})dW_s)^2 = \mathbb{E}\int_0^b \sigma(s, X_s^{(n)})^2 ds \leq \mathbb{E}\int_0^b C(1+(X_s^{(n)})^2)ds < \infty, \tag{4.19}$$

since $X_t^{(n)} \in L^2_{ad}$ by the inductive hypothesis. Since $f(t,x)$ satisfies the linear growth condition as well, we have, using the inductive hypothesis on $X_t^{(n)}$:

$$\mathbb{E}\Big[\int_0^t f(s, X_s^{(n)})ds\Big]^2 \leq \mathbb{E}\Big[\int_0^b C(1+(X_s^{(n)})^2)ds\Big] < \infty. \tag{4.20}$$

Now taking the expectation in (4.18) and inserting (4.19) and (4.20), it follows that also $X_t^{(n+1)} \in L^2$. Since by the inductive hypothesis $X_t^{(n)}$ is adapted, and making a similar observation about the generated σ-algebras as in the case of $X_t^{(1)}$ above, it follows from Lemma 3.28 that $X_t^{(n+1)}$ is adapted as well. Hence by the principle of mathematical induction, we have a sequence of continuous stochastic processes $\{X_t^{(n)}\}_{n\in\mathbb{N}} \subset L^2_{ad}$.

Next, we show that the limit of $\{X_t^{(n)}\}_{n\in\mathbb{N}}$ is also continuous, and solves equation (4.2). For this purpose we begin by estimating $\mathbb{E}(|X_t^{(n+1)} - X_t^{(n)}|^2)$. For notational simplicity, let

$$Y_t^{(n+1)} = \int_0^t \sigma(s, X_s^{(n)})dW_s, \quad Z_t^{(n+1)} = \int_0^t f(s, X_s^{(n)})ds.$$

Now again by $(a+b)^2 \leq 2(a^2+b^2)$ we have:

$$\mathbb{E}(|X_t^{(n+1)} - X_t^{(n)}|^2) \leq 2[\mathbb{E}(|Y_t^{(n+1)} - Y_t^{(n)}|^2) + \mathbb{E}(|Z_t^{(n+1)} - Z_t^{(n)}|^2)]. \tag{4.21}$$

Similarly to what was done earlier when showing the uniqueness of the solution, the stochastic integral term on the right-hand side in the inequality above can be transformed to a Lebesgue integral via the Itô isometry, and then we can use the Lipschitz continuity of $\sigma(t,x)$ to estimate

$$\begin{aligned} \mathbb{E}(|Y_t^{(n+1)} - Y_t^{(n)}|^2) &= \int_0^t \mathbb{E}(|\sigma(s, X_s^{(n)}) - \sigma(s, X_s^{(n-1)})|^2) ds \\ &\leq C^2 \int_0^t \mathbb{E}(|X_s^{(n)} - X_s^{(n-1)}|^2) ds. \end{aligned} \tag{4.22}$$

By Jensen's inequality and the Lipschitz continuity of $f(t,x)$ we have:

$$\begin{aligned} |Z_t^{(n+1)} - Z_t^{(n)}|^2 &= |\int_0^t (f(s, X_s^{(n)} - f(s, X_s^{(n-1)}) ds|^2 \\ &\leq C^2 \int_0^t |X_s^{(n)} - X_s^{(n-1)}|^2 ds. \end{aligned} \tag{4.23}$$

From this it follows that

$$\mathbb{E}|Z_t^{(n+1)} - Z_t^{(n)}|^2 \leq C^2 \int_0^t \mathbb{E}(|X_s^{(n)} - X_s^{(n-1)}|^2) ds, \tag{4.24}$$

where we have used Fubini's theorem to bring the expectation inside the integral, justified by the integrand being non-negative and in L^2. Now putting equations (4.22) and (4.24) into (4.21) we have:

$$\mathbb{E}(|X_t^{(n+1)} - X_t^{(n)}|^2) \leq 2C^2 \int_0^t \mathbb{E}(|X_s^{(n)} - X_s^{(n-1)}|^2) ds. \tag{4.25}$$

On the other hand, a simple computation from the definition of $X_t^{(n)}$ yields together with the Itô isometry, Jensen's inequality and the linear growth condition satisfied by both $f(t,x)$ and $\sigma(t,x)$ that we have the inequality:

$$\mathbb{E}(|X_t^{(2)} - X_t^{(1)}|^2) \leq 2C \int_0^t (1 + \mathbb{E}\xi^2) ds < \infty.$$

Thus by Lemma 4.10 applied to $\theta_{n+1} := \mathbb{E}(|X_t^{(n+1)} - X_t^{(n)}|^2)$ we have:

$$\mathbb{E}(|X_t^{(n+1)} - X_t^{(n)}|^2) \leq (1 + \mathbb{E}\xi^2) \frac{(2C^2)^n t^n}{n!}, \quad n \geq 1. \tag{4.26}$$

Now by the triangle inequality we obviously have the estimate:

$$\sup_{t\in[0,T]} |X_t^{(n+1)} - X_t^{(n)}| \leq \sup_{t\in[0,T]} |Y_t^{(n+1)} - Y_t^{(n)}| + \sup_{t\in[0,T]} |Z_t^{(n+1)} - Z_t^{(n)}|,$$

and this in turn implies that:

$$\{\sup_{t\in[0,T]}|X_t^{(n+1)}-X_t^{(n)}|>\frac{1}{n^2}\}\subset\{\sup_{t\in[0,T]}|Y_t^{(n+1)}-Y_t^{(n)}|>\frac{1}{2n^2}\}$$
$$\cup\{\sup_{t\in[0,T]}|Z_t^{(n+1)}-Z_t^{(n)}|>\frac{1}{2n^2}\}.$$

It follows by monotonicity of the probability measure that we have:

$$P(\{\sup_{t\in[0,T]}|X_t^{(n+1)}-X_t^{(n)}|>\frac{1}{n^2}\})\leq P(\{\sup_{t\in[0,T]}|Y_t^{(n+1)}-Y_t^{(n)}|>\frac{1}{2n^2}\})$$
$$+P(\{\sup_{t\in[0,T]}|Z_t^{(n+1)}-Z_t^{(n)}|>\frac{1}{2n^2}\}). \tag{4.27}$$

Next, we proceed by estimating the two components of the right-hand side of inequality (4.27) separately. First, we observe that since conditional expectation is linear, we easily see that if X and Y are martingales with respect to the same filtration, their difference is as well. Second, since the mapping $x\mapsto|x|$ is convex, we have by Proposition 2.15 that $|X|$ is a submartingale.

By Lemma 3.28 the Itô integral is a martingale, so by the previous discussion, we have that $|Y_t^{(n+1)}-Y_t^{(n)}|$ is a submartingale. Therefore by first applying Lemma 2.20 (Doob's submartingale inequality) in the first term on the right-hand side of inequality (4.27) we can estimate, with the aid of Jensen's inequality:

$$P(\{\sup_{t\in[0,T]}|Y_t^{(n+1)}-Y_t^{(n)}|>\frac{1}{2n^2}\})\leq 2n^2\mathbb{E}(|Y_T^{(n+1)}-Y_T^{(n)}|)$$
$$\leq 4n^4\mathbb{E}(|Y_T^{(n+1)}-Y_T^{(n)}|^2).$$

Continuing the above chain of inequalities by first using the Itô isometry and then the Lipschitz continuity of $\sigma(t,x)$ we obtain:

$$\begin{aligned}P(\{\sup_{t\in[0,T]}|Y_t^{(n+1)}-Y_t^{(n)}|>\frac{1}{2n^2}\}) &= 4n^4\mathbb{E}(|Y_T^{(n+1)}-Y_T^{(n)}|^2)\\ &= 4n^4\int_0^T(\mathbb{E}|\sigma(s,X_s^{(n)})-\sigma(s,X_s^{(n-1)})|^2)ds\\ &\leq 4n^4C^2\int_0^T\mathbb{E}(|X_s^{(n)}-X_s^{(n-1)}|^2)ds\\ &\leq 4n^4C^2(1+\mathbb{E}\xi^2)\frac{(2C^2)^{(n-1)}T^n}{n!},\end{aligned} \tag{4.28}$$

with the ultimate inequality following readily from inequality (4.26) upon integration.

To estimate the second term on the right-hand side of inequality (4.27) we start by first using the Chebyshev inequality:

$$
\begin{aligned}
P(\{\sup_{t\in[0,T]} |Z_t^{(n+1)} - Z_t^{(n)}| > \frac{1}{2n^2}\}) &\leq 4n^4 \mathbb{E}\Big(\sup_{t\in[0,T]} |Z_t^{(n+1)} - Z_t^{(n)}|\Big)^2 \\
&= 4n^4 \mathbb{E}\Big(\sup_{t\in[0,T]} |Z_t^{(n+1)} - Z_t^{(n)}|^2\Big), \qquad (4.29)
\end{aligned}
$$

with the equality following from the elementary fact that for any collection $a_i \subset \mathbb{R}_+$ we have $(\sup a_i)^2 = \sup a_i^2$. Next, note that taking the supremum for $t \in [0,T]$ in inequality (4.23) yields:

$$\sup_{t\in[0,T]} |Z_t^{(n+1)} - Z_t^{(n)}|^2 \leq C^2 \int_0^T |X_s^{(n)} - X_s^{(n-1)}|^2 ds,$$

and continuing this estimate with inequality (4.26) we have, after straightforward integration:

$$\sup_{t\in[0,T]} |Z_t^{(n+1)} - Z_t^{(n)}|^2 \leq C^2(1+\mathbb{E}\xi^2)\frac{(2C^2)^{(n-1)}T^n}{n!}. \qquad (4.30)$$

Now inserting inequality (4.30) to (4.29) we have

$$P(\{\sup_{t\in[0,T]} |Z_t^{(n+1)} - Z_t^{(n)}| > \frac{1}{2n^2}\}) \leq 4n^4C^2(1+\mathbb{E}\xi^2)\frac{(2C^2)^{(n-1)}T^n}{n!}. \qquad (4.31)$$

Finally, inserting equations (4.28) and (4.31) into (4.27) we have

$$P(\{\sup_{t\in[0,T]} |X_t^{(n+1)} - X_t^{(n)}| > \frac{1}{n^2}\}) \leq 8(1+\mathbb{E}\xi^2)\frac{n^4(2C^2)^{n-1}T^n}{n!}.$$

It is well-known that the series on the right-hand side in the above equation converges. Hence it follows from the Borel-Cantelli lemma that we have:

$$P(\{\sup_{t\in[0,T]} |X_t^{(n+1)} - X_t^{(n)}| > \frac{1}{n^2} \quad i.o.\}) = 0,$$

and so in turn for $t \in [0,T]$ the series $\xi + \sum_{n=1}^{\infty}\left(X_t^{(n+1)} - X_t^{(n)}\right)$ converges uniformly for almost every $\omega \in \Omega$ to some limit, denoted by X_t. Note that the nth partial sum of this series is precisely $X_t^{(n)}$. We established earlier that each $X_t^{(n+1)}$ is continuous. Thus as a uniform limit of continuous functions, X_t is continuous. We have shown that $\lim_{n\to\infty} X_t^{(n)} = X_t$ uniformly and almost surely for each $t \in [0,T]$.

What remains is to show that the limit of $\{X_t^{(n)}\}_{n\in\mathbb{N}}$ in fact satisfies equation (4.2) on $[0,T]$. First, note that for any $m,n \in \mathbb{N}$, and say $m > n > 0$, we have, using inequality (4.26):

$$
\begin{aligned}
\sqrt{\mathbb{E}\left|X_t^{(m)} - X_t^{(n)}\right|^2} &= \sqrt{\mathbb{E}\left|\sum_{i=n+1}^{m}(X_t^{(i+1)} - X_t^{(i)})\right|^2} \\
&\le \sum_{i=n+1}^{m}\sqrt{\mathbb{E}\left|X_t^{(i+1)} - X_t^{(i)}\right|^2} \\
&\le \sum_{i=n+1}^{m}\sqrt{(1+\mathbb{E}\xi^2)\frac{(2C^2)^i T^i}{i!}} \\
&\le \sum_{i=n+1}^{\infty}\sqrt{(1+\mathbb{E}\xi^2)\frac{(2C^2)^i T^i}{i!}} \to 0, \quad \text{as } n\to\infty,
\end{aligned}
\tag{4.32}
$$

being a tail of a convergent series. Hence $\{X_t^{(n)}\}_{n\in\mathbb{N}}$ converges in $L^2(\Omega,P)$. Furthermore, since each $\{X_t^{(n)}\}_{n\in\mathbb{N}}$ is adapted, and measurability is preserved under limits, the limit is adapted as well.

By the Itô isometry and the Lipschitz continuity of $\sigma(t,x)$ we have that

$$
\begin{aligned}
\left\| \int_0^t \sigma(s,X_s^{(n)})dW_s - \int_0^t \sigma(s,X_s)dW_s \right\|_{L^2(P)}^2 &= \int_0^t \left\|(\sigma(s,X_s^{(n)}) - \sigma(s,X_s))\right\|_{L^2(P)}^2 ds \\
&\le \int_0^t C\left\|X_s^{(n)} - X_s\right\|_{L^2(P)}^2 ds
\end{aligned}
$$

converges to 0 in $L^2(P)$ as $n\to\infty$, since $X_s^{(n)} \to X_s$ in $L^2(P)$.

Similarly by using the triangle inequality, alongside with the Lipschitz continuity of $f(t,x)$ we have, as $n\to\infty$:

$$
\int_0^t f(s,X_s^{(n)})ds \to \int_0^t f(s,X_s)ds
$$

in $L^2(P)$. Now by the Lebesgue's dominated convergence theorem we can take the limit as $n \to \infty$ in equation (4.17) and the claim is proved.

- The interval of the solution of equation (4.2) can be extended to $[0, \infty)$:

  We have shown above that we have a continuous, unique solution to the stochastic differential equation (4.2) in the sense of Definition 4.1 on $[0, T]$ for any $T \in \mathbb{R}_+$. But for any non-negative real number a we can find an interval of the form $[0, N], N \in \mathbb{N}$ for which $[0, a] \subset [0, N]$ by the Archimedean property of real numbers. Since this holds for all $N \in \mathbb{N}$, the claim is proved, and the proof is finished.

$\square$

Remark 4.33. It is worth noting that we have followed here the 'classical' approach as explained earlier, but the proof given here could be easily modified with use of the results obtained in Section 3.2 with some simplification of the proof. Note that it follows rather easily from inequality (4.26) that the sequence $\{X_t^{(n)}\}_{n\in\mathbb{N}}$ is convergent in $L^2(P)$, and hence the dominated convergence theorem argument can be invoked, proving the existence of a solution. Additionally, the same inequality yields the continuity of the solution as well, via use of Lemma 3.17 and Remark 3.18

Remark 4.34. We mention here briefly that in a more refined terminology, what we have in fact shown in the above proof that the given SDE has a *strong solution*. Note that the solution X_t is constructed with an explicit a priori knowledge of the filtration $\mathcal{F}$, and intuitively X_t at time t depends only on ξ and the information generated by the Brownian motion up until that point. To relax this requirement, we can instead consider the problem of finding a pair of processes X_t and W_t on a filtered space $(\Omega, P, \mathcal{G}_t, \mathcal{G})$ such that for given $f(t,x)$ and $\sigma(t,x)$ the SDE in equation (4.2) holds almost surely. Such X_t is called a *weak solution*. In addition, if any two solutions, strong or weak, have the same law, then we say that *weak uniqueness* holds. For a more thorough general treatment on the subject, see for instance [17] or [18].

Next, we give two examples of stochastic differential equations as an application of Theorem 4.13. In both of these examples we consider the complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,\infty)}, P)$ as in Theorem 4.13.

Example 4.35. (Ornstein-Uhlenbeck). Consider the stochastic differential equation

$$dX_t = \alpha dW_t - \beta X_t dt, \quad X_0 = x_0, \tag{4.36}$$

where $\alpha \in \mathbb{R}, \beta > 0$, and $x_0 \in \mathbb{R}$ with $t \in [0, T]$ for convenience. The solution to this SDE is called the Ornstein-Uhlenbeck process, first proposed by Leonard Ornstein and George E. Uhlenbeck (see, for instance [38]). Ornstein-Uhlenbeck process has many applications in physical sciences and financial mathematics, see [30] or [3]. We can write the given SDE as an integral equation

$$X_t = x_0 + \alpha W_t - \beta \int_0^t X_s ds,$$

so we see that in the notation of Definition 4.1, we have $\sigma(t, X_t) := \alpha$, and $f(t, X_t) := -\beta X_t$. Both of these clearly satisfy Lipschitz and linear growth conditions in X_t. Also as $X_0 \equiv x_0$, a constant, X_0 is trivially in $L^2(P)$, and also is independent of the Brownian motion W. Hence Theorem 4.13 applies, and we have a unique solution X_t satisfying (4.36) on $[0, T]$.
Next we solve (4.36) explicitly, with a formal computation using the Itô formula. Indeed, let $\theta(t, x) = e^{\beta t} x$. Clearly $\theta \in C^2$. Now $\partial_t \theta(t, x) = \beta e^{\beta t} x$, $\partial_x \theta(t, x) = e^{\beta t}$, and $\partial_x^2 \theta(t, x) = 0$. The Itô formula (3.41) yields that

$$\begin{aligned}
&d\theta(t, X_t) = \partial_t \theta(t, X_t) dt + \partial_x \theta(t, X_t) dX_t + \frac{1}{2} \partial_x^2 \theta(t, X_t)(dX_t)^2 \Leftrightarrow \\
&d(e^{\beta t} X_t) = \beta e^{\beta t} X_t dt + e^{\beta} dX_t,
\end{aligned}$$

and putting $dX_t = \alpha dW_t - \beta X_t dt$ from the given SDE we have:

$$\begin{aligned}
d(e^{\beta t} X_t) &= \beta e^{\beta t} X_t dt + e^{\beta t} \big(\alpha dW_t - \beta X_t dt\big) \\
&= \alpha e^{\beta t} dW_t.
\end{aligned}$$

Equivalently in integral form:

$$e^{\beta X_t} X_t = x_0 + \alpha \int_0^t e^{\beta s} dW_s,$$

which can be written as

$$X_t = x_0 e^{-\beta t} + \alpha \int_0^t e^{\beta(s-t)} dW_s.$$

It is worth noting the similarity of the above solution to the solution of a linear ordinary differential equation of the form

$$y'(t) + \beta y(t) = q(t),$$

which for $t \in [0, T]$ is given by

$$y(t) = y(0) e^{-\beta t} + \int_0^t e^{\beta(s-t)} q(s) ds.$$

A common technique to obtain this solution is via the method of an integrating factor. In fact, although the Itô formula shows that the chain rule from ordinary calculus does not carry over to the Itô stochastic calculus unmodified, a similar technique using integrating factors can be developed for solving linear SDE's using the corresponding Leibniz rule for Itô integrals. For more details, see [21].

Example 4.37. (Geometric Brownian motion). Consider the stochastic differential equation

$$dX_t = aX_t dW_t + rX_t dt, \quad X_0 = x_0, \tag{4.38}$$

with $a > 0, r, x_0 \in \mathbb{R}$. The solution to the above equation is called Geometric Brownian motion, and is widely used in financial mathematics, for instance in the Black-Scholes model. Similar to the previous example, we can conclude that Theorem 4.13 is applicable: this time both $f(t, X_t) := aX_t$, and $\sigma(t, x) := rX_t$ are linear in X_t, so the Lipschitz and the linear growth conditions are satisfied, and the condition given for the initial value holds likewise. We solve (4.38) with a formal computation using the Itô formula. Setting $f(x) = \log x$, we have $f'(x) = \frac{1}{x}$, and $f''(x) = -\frac{1}{x^2}$, so $f \notin C^2(\mathbb{R})$, but we proceed formally on $[c, T]$, with $c \neq 0$. Hence by the Itô formula given in (3.39) we have:

$$\begin{aligned} df(X_t) &= f'(X_t)dX_t + \frac{1}{2}f''(X_t)\big(dX_t\big)^2 \Leftrightarrow \\ d\log X_t &= \frac{1}{X_t}dX_t + \frac{1}{2}\Big(-\frac{1}{X_t^2}\Big)\big(dX_t\big)^2. \end{aligned}$$

Inserting $dX_t = aX_t dW_t + rX_t dt$ from the given SDE, and computing $(dX_t)^2$ using the arithmetic $dX_t \cdot dX_t = dt$, $dX_t \cdot dt = dt \cdot dX_t = dt \cdot dt = 0$, we have:

$$\begin{aligned} d\log X_t &= \frac{1}{X_t}\Big(aX_t dW_t - rX_t dt\Big) - \frac{1}{2X_t^2}a^2 X_t^2 dt \\ &= a dW_t + r dt - \frac{1}{2}a^2 dt \\ &= \big(r - \frac{1}{2}a^2\big)dt + a dW_t. \end{aligned}$$

Integrating the above equation over the interval $[c, t]$, we have

$$\log X_t = \log X_c + r(t - c) - \frac{a^2(t - c)}{2} + a(W_t - W_c),$$

from which it follows that the solution to (4.38) is given by:

$$X_t = X_c \exp\Big(t(r - \frac{1}{2}a^2) - c(r - \frac{1}{2}a^2) + a(W_t - W_c)\Big).$$

This can be written as

$$X_t = X_c e^{\alpha t + \beta + a(W_t - W_c)},$$

with $\alpha := r - \frac{1}{2}a^2 \in \mathbb{R}$, and $\beta := -c(r - \frac{1}{2}a^2) \in \mathbb{R}$. Notice in particular that if the initial value for X_0 is non-negative, then the solution X_t will be non-negative for all $t \in [c, T]$ as a consequence.

4.3 Stability of the solution with respect to the initial data

Having established sufficient conditions for an initial value problem (4.13) to be uniquely solvable, we now explore the properties of the solution. A typical question in the ODE theory is to consider the smoothness of the solution with respect to the initial data. Translated to our context, we limit ourselves to the question of continuity of the function $\xi \mapsto X_t(\xi, \omega)$,with ξ being the initial value of the solution. This problem was first considered by Blagovescenskii and Freidlin [4] in 1961.
Our presentation is largely based upon an article by Kunita [19], where the author works in a slightly more general setting by considering the SDE in an n-dimensional space with an additional term given by a Poisson random measure. Also as noted above, we will only prove the continuity of the solution with respect to the initial data under suitable conditions, and for details on other smoothness properties we refer the interested reader to [19]. Finally, it should be noted that the article [19] in turn builds upon an earlier article by Kunita and Fujiwara [9], where the authors work in an even more general setting.
Before proving the main theorem of this section, we need a very useful inequality, which is a part of the Burkholder-Davis-Gundy inequalities, first established in [5]. We will state the following only in the needed generality:

Theorem 4.39. *(Burkholder inequality). Let W be a Brownian motion, and $f \in L^2_{ad}$. Then, for all $p \geq 2$, there exists a constant $C = C(p)$ for which*

$$\mathbb{E}\left[\sup_{t\leq T}\left|\int_0^t f_s dW_s\right|^p\right] \leq C\mathbb{E}\left[\Big(\int_0^T |f(s)|^2 ds\Big)^{\frac{p}{2}}\right], \quad \text{for all } T > 0. \tag{4.40}$$

Proof. Denote $X_t := \int_0^t f_s dW_s$, so that $dX_t = f_t dW_t$. First, note that we can assume that the sample paths of X are almost surely bounded: indeed, define a sequence of stopping times by $\tau_n := \inf\{t \geq 0 : |X_t| = n\}$, and observe that if $T \leq \infty$ the stopped process $X_{\tau_n} \to X_T$, as $n \to \infty$. Hence by a standard monotone convergence argument it is enough to prove the claim in the case that X is bounded. Also, by the non-negativity of the integrand in the right-hand side of inequality (4.40) we can assume that $\mathbb{E}\Big[\sup_{t\le T}\Big|\int_0^t f_s dW_s\Big|^p\Big] \neq 0$.

We first prove the case when $p = 2$. By Lemma 3.33, X_t is a martingale. Hence it follows from Lemma 2.21 and the Itô isometry in Lemma 3.28 that we have

$$\begin{aligned}\mathbb{E}\sup_{t\le T}\left|\int_0^t f_s dW_s\right|^2 &\le 2\sup_{t\le T}\mathbb{E}\left|\int_0^t f_s dW_s\right|^2 = 2\sup_{t\le T}\int_0^t \mathbb{E}\,|f(s)|^2\,ds\\ &\le 2\mathbb{E}\left[\int_0^T |f(s)|^2\,ds\right].\end{aligned}$$

Hence the inequality is proved for $p = 2$.

Next, let $p > 2$. The function $\theta(x) = |x|^p$ is of class C^2 on $\mathbb{R}$. Indeed, a direct computation yields that $\theta'(x) = px|x|^{p-2}$, and $\theta''(x) = p(p-1)|x|^{p-2}$, so we see that $\theta \in C^2(\mathbb{R})$.

Hence we can apply the Itô formula in Theorem 3.40, and obtain from equation (3.41):

$$|X_t|^p = p\int_0^t |X_s|^{p-1} f_t dW_t + \frac{1}{2}p(p-1)\int_0^t |X_s|^{p-2}|f_s|^2 ds,$$

since $(dX_t)^2 = |f_t|^2 dt$. From the above we get immediately

$$\mathbb{E}|X_t|^p = p\mathbb{E}\Big[\int_0^t |X_s|^{p-1} f_t dW_t\Big] + \frac{1}{2}p(p-1)\mathbb{E}\Big[\int_0^t |X_s|^{p-2}|f_s|^2 ds\Big]. \tag{4.41}$$

Since X is bounded, and by assumption $f \in L^2([0,T]\times\Omega, dt\otimes dP)$, a predictable process, the first term in (4.41) is a martingale on $[0,T]$ by Theorem 3.33, and hence has zero expectation. Also by using the boundedness of X we can make the obvious estimate of the second term on the right hand side in the above equation and then we have:

$$\begin{aligned}\mathbb{E}\Big[\int_0^t |X_s|^{p-2}|f_s|^2 ds\Big] &\le \mathbb{E}\Big[\sup_{t\le T}|X_t|^{p-2}\int_0^t |f_s|^2 ds\Big]\\ &\le \Big[\mathbb{E}\sup_{t\le T}|X_t|^p\Big]^{\frac{p-2}{p}}\Big[\mathbb{E}\big(\int_0^T |f_s|^2 ds\big)^{p/2}\Big]^{2/p},\end{aligned} \tag{4.42}$$

where the latter inequality follows from an application of the Hölder inequality with dual exponents $p' = \frac{p}{p-2}$, and $q' = \frac{p}{2}$. Now, as stated before, X_t is a martingale, so from

Theorem 2.21 it follows that we have:

$$\mathbb{E}\sup_{t\le T}|X_t|^p \le \Big(\frac{p}{p-1}\Big)^p \sup_{t\le T}\mathbb{E}|X_t|^p.$$

Hence putting (4.41) into the above inequality and using the estimate in (4.42), and ultimately dividing by the common factor $\mathbb{E}\Big[\sup_{t\le T}|X_t|^p\Big]^{\frac{p-2}{p}}$, we obtain precisely the claimed inequality (4.40), and the proof is finished. □

Remark 4.43. The above inequality remains true if we replace X_t with a local (continuous) martingale. Also the requirement that $p \ge 2$ can be dropped, as in fact the stated inequality holds for any $p > 0$. For details of these claims, see for instance [10] or [7] for the non-continuous case.

Remark 4.44. In Theorem 4.13 we essentially proved that under the given assumptions, the solution of the SDE is in $L^2(P)$. We achieved this with aid of the Itô isometry to convert the stochastic integrals to ordinary Riemann integrals. In fact, the Burkholder inequality allows us to show that for $p \ge 2$, the solution is in fact in $L^p(P)$, assuming that the initial value $\xi \in L^p(P)$. For a proof of this claim, see for instance [19].

Now we are ready to state and prove the main result of this section:

Theorem 4.45. *Let a stochastic differential equation be given by* (4.2)*, with the coefficients $f(t,x), \sigma(t,x)$ satisfying the assumptions of Theorem 4.13. Assume that the initial condition $\xi \equiv x \in \mathbb{R}$ is deterministic. Denote by $X_t(x)$ the solution of* (4.2) *with the initial condition x. Then for any $p \ge 2$, there exists a constant $C = C(p)$, such that*

$$\mathbb{E}\Big[\sup_{s\in[0,t]}|X_s(x) - X_s(y)|^p\Big] \le C|x-y|^p, \quad \forall x, y \in \mathbb{R},$$

holds for any $t \in [0,T]$, with $T > 0$.
In addition, the mapping $x \mapsto X_t(x)$ has a modification which is continuous.

Proof. Note that in order to prove the theorem it is enough to prove the claimed inequality, since it is precisely what is needed to invoke Theorem 2.27 with $\alpha := p > 0, \beta := 1 - p > 0$ to obtain a continuous modification of X_t, proving the latter claim of the theorem.
To prove the inequality, we let $X_t(x)$ and $X_t(y)$ be as defined above, and consider $T > 0$ fixed. Note that we have by the Jensen inequality for $a, b, c \in \mathbb{R} : (a+b+c)^p/3^p \le$

$\frac{a^p+b^p+c^p}{3}$. Hence we have:

$$
\begin{aligned}
|X_t(x)-X_t(y)|^p \leq 3^{p-1}\Big(|x-y|^p + \Big|\int_0^t [b(s,X_s(x))-b(s,X_s(y))]ds\Big|^p \\
+\Big|\int_0^t [\sigma(s,X_s(x))-\sigma(s,X_s(y))]dW_s\Big|^p\Big).
\end{aligned}
\tag{4.46}
$$

We can estimate the second term in (4.46) by using the Hölder inequality with $1/p + 1/q = 1$, and then the Lipschitz condition of $f(t,x)$:

$$
\begin{aligned}
\Big|\int_0^t [b(s,X_s(x))-b(s,X_s(y))]ds\Big|^p &\leq \Big(\int_0^t |b(s,X_s(x))-b(s,X_s(y)|^p ds\Big)^{p/p}\Big(\int_0^t 1^q ds\Big)^{p/q} \\
&= t^{p-1}\int_0^t |b(s,X_s(x))-b(s,X_s(y)|^p ds \\
&\leq C^p t^{p-1}\int_0^t |X_s(x)-X_s(y)|^p ds.
\end{aligned}
$$

Now taking the supremum over $t \leq T$ and then taking the expectation of the above inequality yields:

$$
\begin{aligned}
&\mathbb{E}\sup_{t\leq T}\Big|\int_0^t [b(s,X_s(x))-b(s,X_s(y))]ds\Big|^p \\
&\qquad \leq C^p T^{p-1}\int_0^T \mathbb{E}\sup_{r\leq s}|X_r(x)-X_r(y)|^p ds,
\end{aligned}
\tag{4.47}
$$

where we have used the Fubini Theorem and the Jensen inequality to bring the expectation and the supremum inside the integral.
Next, taking the supremum over $t \leq T$ and the expectation in (4.46), we have for the third term on the right-hand side, by Theorem 4.39:

$$
\mathbb{E}\sup_{t\leq T}\Big|\int_0^t [\sigma(s,X_s(x)-\sigma(s,X_s(y))]dW_s\Big|^p \leq C'\mathbb{E}\Big(\int_0^T \big|\sigma(s,X_s(x))-\sigma(s,X_s(y))\big|^2 ds\Big)^{p/2},
$$

with $C' = C'(p) \in \mathbb{R}_+$. Next, using the Lipschitz continuity of $\sigma(t,x)$ we can continue the above estimate by

$$
C'\mathbb{E}\Big(\int_0^T \big|\sigma(s,X_s(x))-\sigma(s,X_s(y))\big|^2 ds\Big)^{p/2} \leq C'C^p\mathbb{E}\Big[\int_0^T |X_s(x)-X_s(y)|^2 ds\Big]^{p/2}.
$$

Now, assume that $p > 2$. By the Hölder inequality, with $\frac{p}{2}$ and $\frac{p}{p-2}$ being the dual exponents, we have

$$\begin{aligned}\mathbb{E}\Big[\int_0^T |X_s(x) - X_s(y)|^2 ds\Big]^{p/2} &\leq C'C^p\mathbb{E}\Big[\Big(\int_0^T |X_s(x) - X_s(y)|^p ds\Big)^{2/p}\Big(\int_0^T 1^{\frac{p}{p-2}} ds\Big)^{\frac{p-2}{p}}\Big]^{p/2} \\ &= C'CT^{\frac{p-2}{2}}\mathbb{E}\int_0^T |X_s(x) - X_s(y)|^p ds.\end{aligned}$$

Note that this inequality holds trivially in the case $p = 2$, with $C' = C \equiv 1$. By the above chain of inequalities we have, after again using the Jensen inequality to bring the supremum inside the integral, the estimate

$$\begin{aligned}&\mathbb{E}\sup_{t\leq T}\Big|\int_0^t [\sigma(s, X_s(x) - \sigma(s, X_s(y))]dW_s\Big|^p \\ &\qquad\leq C'C^pT^{\frac{p-2}{2}}\int_0^T \mathbb{E}\sup_{r\leq s}|X_r(x) - X_r(y)|^p ds.\end{aligned} \tag{4.48}$$

Now, combining inequalities (4.47) and (4.48) with (4.46) we have, after renaming the constants:

$$\mathbb{E}\sup_{t\leq T}|X_t(x) - X_t(y)|^p \leq C|x - y|^p + C'\int_0^T \mathbb{E}\sup_{r\leq s}|X_r(x) - X_r(y)|^p ds,$$

so applying Lemma 4.8, with $g(T) := \mathbb{E}\sup_{t\leq T}|X_t(x) - X_t(y)|^p, f(T) := C|x-y|^p, \beta := C'$ we have that

$$\mathbb{E}\sup_{t\leq T}|X_t(x) - X_t(y)|^p \leq C|x - y|^p e^{C'T}.$$

This is precisely the inequality that was to be proved. As noted at the beginning of the proof, we can now apply Theorem 2.27, with $\alpha := p > 0, \beta := 1 - p > 0$, to obtain a continuous modification of $x \mapsto X_t(x)$ on $[0, T]$. Since T was arbitrary, by Remarks 4.33 and 4.34 at the end of Theorem 4.13, the claim holds for any $T > 0$, and the proof is finished. □

Chapter 5

Concluding remarks

In this thesis we considered the problem of guaranteeing a unique solution to the stochastic differential equation

$$X_t = \xi + \int_0^t \sigma(s, X_s) dW_s + \int_0^t f(s, X_s) ds \tag{5.1}$$

with σ and f being Lipschitz coefficients of the equation, and ξ a random variable in $L^2(P)$. The integration theory necessary in order to understand the first integral term in equation (5.1) was presented in Chapter 3. Many of the resulting properties of the integral, in particular the Itô isometry, were used in the proof of the existence and uniqueness result for (5.1). It is worth pointing out that generalizing equation (5.1) to higher dimensions is natural. Considering a system of stochastic differential equations is a rather similar task as in the theory of ordinary differential equations, wherein the difficulty lies chiefly in keeping track of notation, as the multidimensional Brownian motion is a rather user-friendly object due to the properties of the normal distribution. A perhaps more fruitful alternative is to consider integration with respect to more general martingales than Brownian motion, as hinted in Remark 3.20. In fact, some parts of Chapter 3 are indeed well-suited for this, although for instance the Itô isometry does not translate verbatim to a more general continuous square-integrable semimartingale M. Instead, we need to consider the quadratic variation process of M stemming from its Doob-Meyer decomposition, and adjust the Itô isometry accordingly. In addition, one could consider the properties of the solution, such as the strong or weak Markov property under similar assumptions as in the main result of Chapter 4.
Finally, we presented the proof of the continuity of the solution with respect to the initial data, when $\xi \in \mathbb{R}$. Additional extension of this is to consider differentiability of the solution with respect to ξ, and other topological properties. This was done already in [9], and later extended to manifolds by Fujiwara in [8].

Bibliography

[1] Ehsan Azmoodeh, Tommi Sottinen, Lauri Viitasaari, and Adil Yazigi. Necessary and sufficient conditions for Hölder continuity of Gaussian processes. *Statistics & Probability Letters*, 94:230 – 235, 2014.

[2] Richard F. Bass. *Diffusions and Elliptic Operators*. Springer, 2006.

[3] Richard F. Bass. *Stochastic Processes*. Cambridge University Press, 2011.

[4] Yuri N. Blagovescenskii and Mark I. Freidlin. Certain properties of diffusion processes depending on parameter. *Sov. Math*, 2:633–636, 1961.

[5] Donald L. Burkholder, Burgess J. Davis, and Richard F. Gundy. Integral inequalities for convex functions of operators on martingales. In *Proceedings of Sixth Berkeley Symposium on Mathematical Statistics and Probability, vol.2*, pages 223–240. University of California Press, 1972.

[6] Erhan Çinlar. *Probability and Stochastics*. Springer, 2010.

[7] Samuel N. Cohen and Robert J. Elliott. *Stochastic Calculus and Applications*. Birkhäuser, 2010.

[8] Tsukasa Fujiwara. Stochastic differential equations of jump type on manifolds and Lévy flows. *J. Math. Kyoto Univ.*, 31(1):99–119, 1991.

[9] Tsukasa Fujiwara and Hiroshi Kunita. Stochastic differential equations of jump type and Lévy processes in diffeomorphisms group. *J. Math Kyoto Univ.*, 25:71–106, 1985.

[10] Jean-François Le Gall. *Brownian Motion, Martingales and Stochastic Calculus*. Springer, 2016.

[11] István Gyöngy and Nikolai V. Krylov. On stochastic equations with respect to semimartingales i. *Stochastics*, 4:1–21, 1980.

[12] Martin Hairer. Solving the KPZ equation. *Ann. Math*, 178:559–664, 2013.

[13] Erika Hausenblas. A note on the Itô formula of stochastic integrals in Banach spaces. *Random Operators and Stochastic Equations*, 14:45–58, 03 2006.

[14] Kiyosi Itô. Stochastic integral. *Proc. Imp. Acad. Tokyo*, 20:519–524, 1944.

[15] Kiyosi Itô. On a stochastic integral equation. *Proc. Japan Acad*, 22:32–35, 1946.

[16] Kiyosi Itô and Shinzo Watanabe. Transformation of Markov processes by multiplicative functionals. *Ann. Inst. Fourier*, 15:15–30, 1965.

[17] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, 1988.

[18] Nikolai V. Krylov and Alexander K. Zvonkin. On strong solutions of stochastic differential equations. *Sel. Math. Sov. I*, 1:19–61, 1980.

[19] Hiroshi Kunita. Stochastic differential equations based on Lévy processes and stochastic flows of diffeomorphisms. In Malempati M. Rao, editor, *Real and Stochastic Analysis. New Perspectives*, pages 305–374. Birkhäuser, 2004.

[20] Hiroshi Kunita and Shinzo Watanabe. On square-integrable martingales. *Nagoya J. Math*, 36:1–26, 1967.

[21] Hui-Hsiung Kuo. *Introduction to Stochastic Integration*. Springer, 2006.

[22] Paul Lévy. Sur certains processus stochastiques homogènes. *Compositio Mathematica*, 7:283–339, 1940.

[23] Paul Lévy. *Processus Stochastiques et Mouvement Brownien*. Gauthier-Villars, 1948.

[24] Robert S. Liptser and Albert N. Shiryayev. *Statistics of random processes I, General theory*. Springer, 1999.

[25] Terry J. Lyons. Differential equations driven by rough signals. *Revista matemática iberoamericana*, 14:215–310, 1998.

[26] Vidyadhar Mandrekar and B. Rüdiger. Existence and uniqueness of path wise solutions for stochastic integral equations driven by Lévy noise on separable Banach spaces. *Stochastics: An International Journal of Probability and Stochastics Processes*, 78:189–212, 08 2006.

[27] Paul-André Meyer. Integrales stochastiques. *Lecture Notes in Mathematics*, 39:72–162, 1967.

[28] David Nualart. *The Malliavin Calculus and Related Topics*. Springer, 2006.

[29] Marcel Nutz. Pathwise construction of stochastic integrals. *Electron. Commun. Probab.*, 17:1–7, 2012.

[30] Berndt Øksendal. *Stochastic Differential Equations*. Springer, 2000.

[31] Etienne Pardoux and Aurel Răşcanu. *Stochastic Differential Equations, Backward SDEs, Partial Differential Equations*. Springer, 2014.

[32] Philip Protter. *Stochastic Integration and Differential Equations*. Springer, 2005.

[33] Norbert Wiener Raymond Paley and Antoni Zygmund. Notes on random functions. *Math. Zietschr.*, 37:647–668, 1933.

[34] Daniel Revuz and Marc Yor. *Continuous Martingales and Brownian Motion*. Springer, 2005.

[35] René L. Schilling and Lothar Partzsch. *Brownian Motion. An Introduction to Stochastic Processes*. De Gruyter, 2012.

[36] Ruslan Stratonovich. A new representation for stochastic integrals and equations. *J. Siam Control*, 4:362–371, 1966.

[37] Daniel W. Stroock and S.R. Srnivasa Varadhan. *Multidimensional Diffusion Processes*. Springer, 1997.

[38] Leonard Uhlenbeck and George E. Ornstein. On the theory of the Brownian motion. *Physical Review*, 36:823–841, 1930.

[39] Norbert Wiener. Differential space. *J. Math. Phys.*, 58:131–174, 1923.

[40] Eugene Wong and Moshe Zakai. On the convergence of ordinary integrals to stochastic integrals. *Ann. Math. Stat.*, 36:1560–1564, 1965.

[41] James Yeh. *Martingales and Stochastic Analysis*. World Scientific, 1995.

[42] James Yeh. *Real Analysis. Theory of Measure and Integration*. World Scientific, 2006.

[43] Jianfeng Zhang. *Backward Stochastic Differential Equations*. Springer, 2007.